Quantum Theory over a Galois Field and Applications to Gravity and Particle Theory

Felix M. Lev

Artwork Conversion Software Inc., 1201 Morningside Drive, Manhattan Beach, CA 90266, USA (Email: felixlev314@gmail.com)

Abstract:

We argue that the main reason of crisis in quantum physics is that nature, which is fundamentally discrete, is described by continuous mathematics. Moreover, no ultimate physical theory can be based on continuous mathematics because, as follows from Gödel's incompleteness theorems, that mathematics is not self-consistent. In the first part of the work we discuss inconsistencies in standard approach to quantum theory and reformulate the theory such that it can be naturally generalized to a formulation based on discrete mathematics. It is shown that the cosmological acceleration and gravity can be treated simply as *kinematical* manifestations of de Sitter symmetry on quantum level (*i.e. for describing those phenomena the notions of dark energy, space-time background and gravitational interaction are not needed*). In the second part of the work we argue that fundamental quantum theory should be based on a Galois field with a large characteristic p. In this approach the de Sitter gravitational constant depends on p and disappears in the formal limit $p \to \infty$, i.e. gravity is a consequence of finiteness of nature. The application of the approach to particle theory gives the following results: a) no neutral elementary particles can exist; b) the electric charge and the baryon and lepton quantum numbers can be only approximately conserved (i.e. the notion of a particle and its antiparticle is only approximate). We also consider a possibility that only Dirac singletons can be true elementary particles.

PACS: 02.10.Hh, 11.30.Fs, 11.30.Ly, **12.90.+b**

Keywords: quantum theory, Galois fields, de Sitter invariance, gravity

Contents

Chapter 1

Introduction

1.1 What is the main reason of crisis in physics?

The discovery of atoms and elementary particles indicates that at the very fundamental level nature is discrete. As a consequence, any description of macroscopic phenomena by continuous mathematics can be only approximate. For example, the water in the ocean can be described by equations of hydrodynamics but we know that this is only an approximation since matter is discrete. Analogously, the notion of continuous geometry has originated from macroscopic experience but this geometry cannot be used for describing physics at the fundamental level. It is also obvious that standard division and the notion of infinitely small are based on our everyday experience that any macroscopic object can be divided by two, three and even a million parts. But is it possible to divide by two or three the electron or neutrino? It seems obvious that the very existence of elementary particles indicates that standard division has only a limited meaning. Indeed, consider, for example, the gram-molecule of water having the mass 18 grams. It contains the Avogadro number of molecules $6 \cdot 10^{23}$. We can divide this gram-molecule by ten, million, billion, but when we begin to divide by numbers greater than the Avogadro one, the division operation loses its meaning.

It is interesting to note that even the name "quantum theory" reflects a belief that nature is quantized, i.e. discrete. Nevertheless, when quantum theory was created it was based on continuous mathematics developed mainly in the 19th century when people did not know about atoms and elementary particles and believed that every macroscopic object could be divided by any number of parts. One of the greatest successes of the early quantum theory was the discovery that energy levels of the hydrogen atom can be described in the framework of continuous mathematics because the Schrödinger differential operator has a discrete spectrum. This and many other successes of quantum theory were treated as indications that all problems of the theory can be solved by using continuous mathematics.

As a consequence, even after almost 90 years of the existence of quantum theory it is still based on continuous mathematics. Although the theory contains divergencies and other inconsistencies, physicists persistently try to resolve them in the framework of continuous mathematics. For example, many physicists believe that M theory or string theory will become the "theory of everything". In those theories physics depends on topology at Planck distances ($10^{-35}m$). Meanwhile the lessons of quantum theory tell us that at such distances (and even much greater ones) no continuous topology or geometry can describe physics.

It is also very important to note that even continuous mathematics by itself has its own foundational problems. Indeed, as follows from Gödel's incompleteness theorems, no system of axioms can ensure that all facts about natural numbers can be proved. Moreover, the system of axioms in standard mathematics cannot demonstrate its own consistency. The theorems demonstrate that any mathematics involving the set of all natural numbers is not self-consistent. Therefore one might expect that the ultimate quantum theory will be based on mathematics which is not only discrete but even finite. Additional arguments in favor of this statement are given in Sec. 6.1.

The reason why modern quantum physics is based on continuity, differentiability etc. is probably historical: although the founders of quantum theory and many physicists who contributed to it were highly educated scientists, discrete mathematics was not (and still is not) a part of standard physics education.

The main problem is the choice of strategy for constructing a new quantum theory. Since no one knows for sure what strategy is the best one, different approaches should be investigated. Dirac's advice given in Ref. [1] is: *"I learned to distrust all physical concepts as a basis for a theory. Instead one should put one's trust in a mathematical scheme, even if the scheme does not appear at first sight to be connected with physics. One should concentrate on getting an interesting mathematics."*

I understand this advice such that our macroscopic experience and physical intuition do not work on quantum level and hence here we can rely only on solid mathematics. However, many physicists do not think so and believe that Dirac was "The Strangest Man" (this is the title of the book by Graham Farmelo about Dirac).

In view of the above remarks and Dirac's advice it seems natural that fundamental quantum physics should be based on discrete mathematics. Beginning from Chap. 6 we consider an approach where quantum theory is based on a Galois field rather than the field of complex numbers. At the same time, one of the key principles of physics is the correspondence principle. It means that at some conditions any new theory should reproduce results of the old well tested theory with a good accuracy. Usually the correspondence principle is applied such that the new theory contains a parameter and reproduces results of the old theory in a formal limit when the parameter is infinitely large or infinitely small. Well-known examples are that nonrelativistic theory is a special case of relativistic one in the formal limit $c \to \infty$ and classical (i.e. nonquantum) theory is a special case of quantum one in the formal limit $\hbar \to 0$ (see however a discussion in Sec. 1.4).

Hence one should find a formulation of standard continuous physics which can be naturally generalized to a formulation based on discrete mathematics. This problem is discussed in the first part of this work. Beginning from Chap. 6 we consider a quantum theory over a Galois field (GFQT) which is not only discrete but even finite. In particular, GFQT does not contain infinitely small and infinitely large quantities and here divergencies cannot exist in principle since any Galois field is finite. Standard theory can be treated as a special case of GFQT in a formal limit $p \to \infty$ where p is the characteristic of the Galois field in GFQT.

1.2 Does quantum theory need space-time?

The phenomenon of quantum field theory (QFT) has no analogs in the history of science. There is no branch of science where so impressive agreements between theory and experiment have been achieved. At the same time, the level of mathematical rigor in QFT is very poor and, as a result, QFT has several well-known difficulties and inconsistencies. The absolute majority of physicists believe that agreement with experiment is much more important than the lack of mathematical rigor, but not all of them think so. For example, Dirac wrote in Ref. [1]: *"The agreement with observation is presumably by coincidence, just like the original calculation of the hydrogen spectrum with Bohr orbits. Such coincidences are no reason for turning a blind eye to the faults of the theory. Quantum electrodynamics is rather like Klein-Gordon equation. It was built up from physical ideas that were not correctly incorporated into the theory and it has no sound mathematical foundation."* In addition, QFT fails in quantizing gravity since the gravitational constant has the dimension (length)^2 (in units where $c = \hbar = 1$), and as a result, quantum gravity is not renormalizable.

Usually there is no need to require that the level of mathematical rigor in physics should be the same as in mathematics. However physicists should have a feeling that, at least in principle, mathematical statements used in the theory can be substantiated. The absence of a well-substantiated QFT by no means can be treated as a pure academic problem. This becomes immediately clear when one wants to work beyond perturbation theory. The problem arises to what extent the difficulties of QFT can be resolved in the framework of QFT itself or QFT can only be a special case of a more general theory based on essentially new ideas. The majority of physicists believe that QFT should be treated [2] *"in the way it is"*, but at the same time it is [2] a *"low energy approximation to a deeper theory that may not even be a field theory, but something different like a string theory"*.

One of the key ingredients of QFT is the notion of space-time background. We will discuss this notion in view of the measurability principle, i.e. that a definition of a physical quantity is a description of how this quantity should be measured. In particular, the Copenhagen interpretation is based on this principle. In this interpretation the process of measurement necessarily implies an interaction with a classical object. This interpretation cannot be universal since it does not consider situations when the world does not have classical objects at all. Meanwhile in cosmological theories there were no classical objects at the early stages of the world. The problem of interpretation of quantum theory is still open but it is commonly accepted that at least at the present stage of the world the measurability principle is valid.

Since physics is based on mathematics, intermediate stages of physical theories can involve abstract mathematical notions but any physical theory should formulate its final results only in terms of physical (i.e. measurable) quantities. Typically the theory does not say explicitly how the physical quantities in question should be measured (a well-known exclusions are special and general theories of relativity where the distances should be measured by using light signals) but it is assumed that in principle the measurements can be performed. In classical (i.e. nonquantum) theory it is assumed that any physical quantity in the theory can be measured with any desired accuracy. In quantum theory the measurability principle is implemented by requiring that any physical quantity can be discussed only in conjunction with an operator defining this quantity. However, quantum theory does not specify how the operator of a physical quantity is related to the measurement of this quantity.

In classical nonrelativistic mechanics, the space-time background is the four-dimensional Galilei space, the coordinates (t, x, y, z) of which are in the range $(-\infty, \infty)$. The set of all points of Galilei space is treated as *a set of possible events for real particles in question* and the assumption is that at each moment of time t the spatial coordinates (x, y, z) of any particle can be measured with the absolute accuracy. Then a very important observation is that, from the point of view of the measurability principle, Galilei space has a physical meaning only as a *space of events for real particles* while if particles are absent, the notion of empty Galilei space has no physical meaning. Indeed, there is no way to measure coordinates of a space which exists only in our imagination. In mathematics one can use different spaces regardless of whether they have a physical meaning or not. However, in physics spaces which have no physical meaning can be used only at intermediate stages. Since in classical mechanics the final results are formulated in terms of Galilei space, this space should be physical.

In classical relativistic mechanics, the space-time background is the four-dimensional Minkowski space and the above remarks can be applied to this space as well. The distances in Minkowski space are defined by the diagonal metric tensor $\eta_{\mu\nu}$ such that $\mu, \nu = 0, 1, 2, 3$ and $\eta_{00} = -\eta_{11} = -\eta_{22} = -\eta_{33} = 1$. Minkowski space is also the space-time background in classical electrodynamics. Here the Maxwell equations make it possible to calculate the electric and magnetic fields, $\mathbf{E}(t, x, y, z)$ and $\mathbf{B}(t, x, y, z)$, at each point of Minkowski space. These fields can be measured by using test bodies at different moments of time and different positions. Hence in classical electrodynamics, Minkowski space can be physical only in the presence of test bodies but not as an empty space.

In General Relativity (GR) the range of the coordinates (t, x, y, z) and the geometry of space-time are dynamical. They are defined by the Einstein equations

$$R_{\mu\nu} + \frac{1}{2}g_{\mu\nu}R_c + \Lambda g_{\mu\nu} = (8\pi G/c^4)T_{\mu\nu} \qquad (1.1)$$

where $R_{\mu\nu}$ is the Ricci tensor, R_c is the scalar curvature, $T_{\mu\nu}$ is the stress-energy tensor of matter, $g_{\mu\nu}$ is the metric tensor, G is the gravitational constant and Λ is the cosmological constant (CC). In modern quantum theory space-time in GR is treated as a description of quantum gravitational field in classical limit. On quantum level each field is a collection of particles; in particular it is believed that the gravitational field is a collection of gravitons. From this point of view the following question arises. Why does $T_{\mu\nu}$ describe the contribution of electrons, protons, photons and other particles but gravitons are not included into $T_{\mu\nu}$ and are described separately by a quantized version of $R_{\mu\nu}$? In any case, quantum theory of gravity has not been constructed yet and gravity is known only at macroscopic level. Here the coordinates and the curvature of space-time are the physical quantities since the information about them can be obtained from measurements using (macroscopic) test bodies. Since matter is treated as a source of the gravitational field, in the formal limit when matter disappears, the gravitational field should disappear too. Meanwhile, in this limit the solutions of Eq. (1.1) are Minkowski space when $\Lambda = 0$, de Sitter (dS) space when $\Lambda > 0$ and anti-de Sitter (AdS) space when $\Lambda < 0$. Hence in GR, Minkowski, dS or AdS spaces can be only empty spaces, i.e. they are not physical. This shows that the formal limit of GR when matter disappears is nonphysical since in this limit the space-time background survives and has a curvature - zero curvature in the case of Minkowski space and a nonzero curvature in the case of dS or AdS spaces.

To avoid this problem one might try to treat the space-time background as a reference frame. In standard textbooks (see e.g., Ref. [3]) the reference frame in GR is defined as a collection of weightless bodies, each of which is characterized by three numbers (coordinates) and is supplied by a clock. Such a notion (which resembles ether) is not physical even on classical level and for sure it is meaningless on quantum level. In some approaches (see e.g. Ref. [4]), when matter disappears, the metric tensor becomes not the Minkowskian one but zero, i.e. the space-time background disappears too. Also, as argued in Ref. [5], the metric tensor should be dimensionful since $g_{\mu\nu}dx^\mu dx^\nu$ should be scale independent. In this approach the absolute value of the metric tensor is proportional to the number of particles in the World.

In approaches based on holographic principle it is stated that the space-time background is not fundamental but emergent. For example, as noted in Ref. [6], "*Space is in the first place a device introduced to describe the positions and movements of particles. Space is therefore literally just a storage space for information...*". This implies that the emergent space-time background is meaningful only if matter is present. The author of Ref. [6] states that in his approach one can recover Einstein equations where the coordinates and curvature refer to the emergent space-time.

However, it is not clear how to treat the fact that the formal limit when matter disappears is possible and the space-time background formally remains although, if it is emergent, it cannot exist without matter.

As noted above, from the point of view of quantum theory, any physical quantity can be discussed only in conjunction with an operator defining this quantity. As noted by Pauli (see p. 63 of Ref. [7]), at early stages of quantum theory some authors treated time t as an operator commuting with the Hamiltonian as $[H,t] = i\hbar$. However, such a treatment is not correct. For example, one cannot construct the eigenstate of the time operator with the eigenvalue 5000 BC or 2014 AD. It is usually assumed that in quantum theory the quantity t can be only a classical parameter describing evolution of a quantum system by the time dependent Schrödinger equation (see e.g. Refs. [7, 8]). This poses a problem why the principle of quantum theory that every physical quantity is defined by an operator does not apply to time.

As noted by several authors, (see e.g. Refs. [9, 10]), t cannot be treated as a fundamental physical quantity. The reason is that all fundamental physical laws do not require time and the quantity t is obsolete on fundamental level. A hypothesis that time is an independently flowing fundamental continuous quantity has been first proposed by Newton. However, a problem arises whether this hypothesis is compatible with the principle that the definition of a physical quantity is a description of how this quantity can be measured.

Consider first the problem of time in classical mechanics. A standard treatment of this theory is that its goal is to solve equations of motion and get classical trajectories where coordinates and momenta are functions of t. In Hamiltonian mechanics the action can be written as $S = S_0 - \int H dt$ where S_0 does not depend on t and is called the abbreviated action. Then, as explained in textbooks, the dependence of the coordinates and momenta on t can be obtained from a variational principle with the action S. Suppose now that one wishes to consider a problem which is usually treated as less general: to find not the dependence of the coordinates and momenta on t but only possible forms of trajectories in the phase space without mentioning time at all. If the energy is a conserved physical quantity then, as described in textbooks, this problem can be solved by using the Maupertuis principle involving only S_0.

However, the latter problem *is not* less general than the former one. For illustration we first consider the one-body case. Suppose that by using the Maupertuis principle one has solved the problem with some initial values of coordinates and momenta. Let s be a parameter characterizing the particle trajectory, i.e. the particle radius-vector $\mathbf{r}$, the momentum $\mathbf{p}$ and the energy E are functions of s. The particle velocity $\mathbf{v}$ in units $c = 1$ is defined as $\mathbf{v}(s) = \mathbf{p}(s)/E(s)$. At this stage the problem does not contain t yet. One can *define* t by the condition that $dt = |d\mathbf{r}|/|\mathbf{v}|$ and hence the value of t at any point of the trajectory can be obtained by integration. In the case of many bodies one can define t by using the spatial trajectory of any body and the result does not depend on the choice of the body. Hence the general problem of classical mechanics can be formulated without mentioning t while if for some reasons one prefers to work with t then its value can flow only in the positive direction since $dt > 0$.

In this work we will consider only the case of free particles. Then, as shown in Sec. 5.7, classical equations of motion can be obtained even without using variational principles, Hamilton equations etc. Namely, equations of motion can be derived by using conservation laws and assuming that t is defined such that the coordinates and momenta of each particle are related to each other such that

$$d\mathbf{r} = \mathbf{v}dt = \frac{\mathbf{p}}{E}dt \tag{1.2}$$

where $E = (m^2 + \mathbf{p}^2)^{1/2}$ and m is the particle mass.

Consider now the problem of time in quantum theory. In the case of one strongly quantum system (i.e. the system which cannot be described in classical theory) a problem arises whether there exists a quantum analog of the Maupertuis principle and whether time can be defined by using this analog. This is a difficult unsolved problem. A possible approach for solving this problem has been proposed in Ref. [9]. However, one can consider a situation when a quantum system under consideration is a small subsystem of a big system where the other subsystem - the environment, is strongly classical. Then one can define t for the environment as described above. The author of Ref. [10] considers a scenario when the system as a whole is described by the stationary Schrödinger equation $H\Psi = E\Psi$ but the small quantum subsystem is described by the time dependent Schrödinger equation where t is defined for the environment as $t = \partial S_0/\partial E$.

One might think that this scenario gives a natural solution of the problem of time in quantum theory. Indeed, in this scenario it is clear why a quantum system is described by the Schrödinger equation depending on the classical parameter t which is not an operator: because t is the physical quantity characterizing not the quantum system but the environment. This scenario seems also natural because it is in the spirit of the Copenhagen interpretation of quantum theory: the evolution of a quantum system can be characterized only in terms of measurements which in the Copenhagen interpretation are treated as interactions with classical objects. However, this scenario encounters several problems. For example, the environment can be a classical object only in some approximation and hence t can be only an approximately continuous parameter. In addition, as noted above, the Copenhagen interpretation cannot be universal in all situations.

As noted in Ref. [10], the above scenario also does not solve the problem of quantum jumps. For illustration, consider a photon emitted in the famous 21cm transition line between the hyperfine energy levels of the hydrogen atom. The phrase that the lifetime of this transition is of the order of $\tau = 10^7$ years should be understood such that the width of the level is of the order of $\hbar/\tau$ i.e. the uncertainty of the photon energy is $\hbar/\tau$. In this situation a description of the system (atom + electric field) by the wave function (e.g. in the Fock space) depending on a continuous parameter t has no physical meaning (since roughly speaking the quantum of time in this process is of the order of 10^7 years). If we accept this explanation then we should acknowledge that in some situations a description of evolution by a continuous classical parameter t is not physical. This is in the spirit of the Heisenberg S-matrix program that in quantum theory one can describe only transitions of states from the infinite past when $t \to -\infty$ to the distant future when $t \to +\infty$.

While no operator can be associated with time, a problem arises whether it is possible to consistently define the position operator. This problem is discussed in detail in Chap. 2. However, QFT operates not with position operators for each particle but with local quantum fields. A non-quantized quantum field $\psi(x) = \psi(t, \mathbf{x})$ combines together two irreducible representations (IRs) with positive and negative energies. The IR with the positive energy is associated with a particle and the IR with the negative energy is associated with the corresponding antiparticle. From mathematical point of view, a local quantum field is described by a reducible representation induced not from the little algebra IRs are induced from but from the Lorenz algebra. The local fields depend on x because the factor space of the Poincare group over the Lorentz group is Minkowski space. In that case there is no physical operator corresponding to x, i.e. x is not measurable. Since the fields describe nonunitary representations, their probabilistic interpretation is problematic. In addition, as it has been shown for the first time by Pauli [11] (see also textbooks on QFT, e.g. Chap. 2 in Ref. [12]), in the case of fields with an integer spin it is not possible to construct a positive definite charge operator and in the case of fields with a half-integer spin it is not possible to construct a positive definite energy operator. It is also known that the description of the electron in the external field by the Dirac spinor is not accurate (e.g. it does not take into account the Lamb shift).

Hence a problem arises why we need local fields at all. They are not needed if we consider only systems of noninteracting particles. Indeed, such systems are described by tensor products of IRs and all the operators of such tensor products are well defined. Local fields are used for constructing interacting Lagrangians which in turn, after quantization, define the representation operators of the Poincare algebra for a system of interacting particles under consideration. Hence local fields do not have a direct physical meaning but are only auxiliary notions.

It is known (see e.g. the textbook [13]) that quantum interacting local fields can be treated only as operatorial distributions. A well-known fact from the theory of distributions is that their products at the same point are poorly defined. Hence if $\psi_1(x)$ and $\psi_2(x)$ are two local operatorial fields then the product $\psi_1(x)\psi_2(x)$ is not well defined. This is known as the problem of constructing composite operators. A typical approach discussed in the literature is that the arguments of the field operators ψ_1 and ψ_2 should be slightly separated and the limit when the separation goes to zero should be taken only at the final stage of calculations. However, no universal way of separating the arguments is known and it is not clear whether any separation can resolve the problems of QFT. Physicists often ignore this problem and use such products to preserve locality (although the operator of the quantity x does not exist). As a consequence, the representation operators of interacting systems constructed in QFT are not well defined and the theory contains anomalies and infinities. Also, one of the known results in QFT is the Haag theorem and its generalizations (see e.g. Ref. [14]) that the interaction picture in QFT does not exist. We believe it is rather unethical that even in almost all textbooks on QFT this theorem is not mentioned at all.

While in renormalizable theories the problem of infinities can be somehow circumvented at the level of perturbation theory, in quantum gravity infinities cannot be excluded even in lowest orders of perturbation theory. One of the ideas of the string theory is that if products of fields at the same points (zero-dimensional objects) are replaced by products where the arguments of the fields belong to strings (one-dimensional objects) then there is hope that infinities will be less singular. However, the problem of infinities in the string theory has not been solved yet. As noted above, in spite of such mathematical problems, QFT is very popular since it has achieved great successes in describing many experimental data.

In quantum theory, if we have a system of particles, its wave function (represented as a Fock state or in other forms) gives the maximum possible information about this system and there is no other way of obtaining any information about the system except from its wave function. So if one works with the emergent space, the information encoded in this space should be somehow extracted from the system wave function. However, to the best of our knowledge, there is no theory relating the emergent space with the system wave function. Typically the emergent space is described in the same way as the "fundamental" space, i.e. as a manifold and it is not clear how the points of this manifold are related to the wave function. The above arguments showing that the "fundamental" space is not physical can be applied to the emergent space as well. In particular, the coordinates of the emergent space are not measurable and it is not clear what is the meaning of those coordinates where there are no particles at all.

In Loop Quantum Gravity (LQG), space-time is treated on quantum level as a special state of quantum gravitational field (see e.g. Ref. [15]). This construction is rather complicated and one of its main goals is to have a quantum generalization of space-time such that GR should be recovered as a classical limit of quantum theory. However, so far LQG has not succeeded in proving that GR is a special case of LQG in classical limit.

In view of this discussion, it is unrealistic to expect that successful quantum theory of gravity will be based on quantization of GR or on emergent space-time. The results of GR might follow from quantum theory of gravity only in situations when space-time coordinates of *real bodies* is a good approximation while in general the formulation of quantum theory should not involve the space-time background at all. One might take objection that coordinates of space-time background in GR can be treated only as parameters defining possible gauge transformations while final physical results do not depend on these coordinates. Analogously, although the quantity x in the Lagrangian density $L(x)$ is not measurable, it is only an auxiliary tool for deriving equations of motion in classical theory and constructing Hilbert spaces and operators in quantum theory. After this construction has been done, one can safely forget about background coordinates and Lagrangian. In other words, a problem is whether nonphysical quantities can be present at intermediate stages of physical theories. This problem has a long history discussed in a vast literature. Probably Newton was the first who introduced the notion of space-time background but, as noted in a paper in Wikipedia, "Leibniz thought instead that space was a collection of relations between objects, given by their distance and direction from one another". As noted above, the assumption that space-rime exists and has a curvature even when matter is absent is not physical. We believe that at the fundamental level unphysical notions should not be present even at intermediate stages. So Lagrangian can be at best treated as a hint for constructing a fundamental theory. As stated in Ref. [16], local quantum fields and Lagrangians are rudimentary notion, which will disappear in the ultimate quantum theory. Those ideas have much in common with the Heisenberg S-matrix program and were rather popular till the beginning of the 1970s. In view of successes of gauge theories they have become almost forgotten.

In summary, *there are no physical arguments showing that the notions of space-time background and local quantum fields are needed in quantum theory.* On the other hand, since this theory is treated as more general than the classical one, in quantum theory it is not possible to fully avoid space-time description of real bodies in semiclassical approximation. Indeed, quantum theory should explain how photons from distant stars travel to the Earth and even how one can recover the motion of macroscopic bodies along classical trajectories (see Chap. 2 for a more detailed discussion).

Let us make a few remarks about the terminology of quantum theory. The terms "wave function" and "particle-wave duality" have arisen at the beginning of quantum era in efforts to explain quantum behavior in terms of classical waves but now it is clear that no such explanation exists. The notion of wave is purely classical; it has a physical meaning only as a way of describing systems of many particles by their mean characteristics. In particular, such notions as frequency and wave length can be applied only to classical waves, i.e. to systems consisting of many particles. If a particle state vector contains $exp[i(px-Et)/\hbar]$ then by analogy with the theory of classical waves one might say that the particle is a wave with the frequency $\omega = E/\hbar$ and the (de Broglie) wave length $\lambda = 2\pi\hbar/p$. However, such defined quantities ω and λ are not real frequencies and wave lengths measured e.g. in spectroscopic experiments. The term "wave function" might be misleading since in quantum theory it defines not amplitudes of waves but only amplitudes of probabilities. So, although in our opinion the term "state vector" is more pertinent than "wave function" we will use the latter in accordance with the usual terminology, and the phrase that a photon has a frequency ω and the wave length λ will be understood only such that $\omega = E/\hbar$ and $\lambda = 2\pi\hbar/p$.

In classical theory the notion of field, as well as that of wave, is used for describing systems of many particles by their mean characteristics. For example, the electromagnetic field consists of many photons. In classical theory each photon is not described individually but the field as a whole is described by the quantities $\mathbf{E}(x)$ and $\mathbf{B}(x)$ which, as noted above, can be measured (in principle) by using macroscopic test bodies. However, QFT is based on quantized field operators $\psi(x)$ which contain the information about the state vector of every particle and, as noted above, there is no well defined operator of the four-vector x. In particular, the notions of electric and magnetic fields of an elementary particle have no physical meaning. In view of these observations and the above remarks about quantum fields we believe that the term "quantum field", as well as the term "wave function" might be misleading.

1.3 Symmetry on quantum level

In relativistic quantum theory the usual approach to symmetry on quantum level is as follows. Since Poincare group is the group of motions of Minkowski space, quantum states should be described by representations of the Poincare group. In turn, this implies that the representation generators should commute according to the commutation relations of the Poincare group Lie algebra:

$$\begin{aligned}
&[P^\mu, P^\nu] = 0 \quad [P^\mu, M^{\nu\rho}] = -i(\eta^{\mu\rho}P^\nu - \eta^{\mu\nu}P^\rho) \\
&[M^{\mu\nu}, M^{\rho\sigma}] = -i(\eta^{\mu\rho}M^{\nu\sigma} + \eta^{\nu\sigma}M^{\mu\rho} - \eta^{\mu\sigma}M^{\nu\rho} - \eta^{\nu\rho}M^{\mu\sigma})
\end{aligned} \tag{1.3}$$

where P^μ are the operators of the four-momentum and $M^{\mu\nu}$ are the operators of Lorentz angular momenta. This approach is in the spirit of Klein's Erlangen program in mathematics. However, as we argue in Refs. [17, 18] and in the preceding section, quantum theory should not be based on classical space-time background and the approach should be the opposite. Each system is described by a set of independent operators. By definition, the rules how these operators commute with each other define the symmetry algebra. In particular, *by definition*, Poincare symmetry on quantum level means that the operators commute according to Eq. (1.3). This definition does not involve Minkowski space at all. Such a definition of symmetry on quantum level is in the spirit of Dirac's paper [19].

For understanding this definition the following example might be useful. If we define how the energy should be measured (e.g., the energy of bound states, kinetic energy *etc.*), we have a full knowledge about the Hamiltonian of our system. In particular, we know how the Hamiltonian commutes with other operators. In standard theory the Hamiltonian is also interpreted as an operator responsible for evolution in time, which is considered as a classical macroscopic parameter (see the preceding section). In situations when this parameter is a good approximate parameter, macroscopic transformations from the symmetry group corresponding to the evolution in time have a meaning of evolution transformations. However, there is no guaranty that such an interpretation is always valid (e.g., at the very early stage of the World or in the example with the 21cm transition line discussed in the preceding section). In general, according to principles of quantum theory, self-adjoint operators in Hilbert spaces represent observables but there is no requirement that parameters defining a family of unitary transformations generated by a self-adjoint operator are eigenvalues of another self-adjoint operator. A well-known example from standard quantum mechanics is that if P_x is the x component of the momentum operator then the family of unitary transformations generated by P_x is $exp(iP_x x/\hbar)$ where $x \in (-\infty, \infty)$ and such parameters can be identified with the spectrum of the position operator. At the same time, the family of unitary transformations generated by the Hamiltonian H is $exp(-iHt/\hbar)$ where $t \in (-\infty, \infty)$ and those parameters cannot be identified with a spectrum of a self-adjoint operator on the Hilbert space of our system. In the relativistic case the parameters x can be formally identified with the spectrum of the Newton-Wigner position operator [20] but, as noted in the preceding section and shown in Chap. 2, this operator does not have all the required properties for the position operator. So, although the operators $exp(iP_x x/\hbar)$ and $exp(-iHt/\hbar)$ are formally well defined, their physical interpretation as translations in space and time is questionable.

Analogously, the definition of the dS symmetry on quantum level should not involve the fact that the dS group is the group of motions of the dS space. Instead, *the definition* is that the operators M^{ab} ($a, b = 0, 1, 2, 3, 4$, $M^{ab} = -M^{ba}$) describing the system under consideration satisfy the commutation relations *of the dS Lie algebra* so(1,4), *i.e.*,

$$[M^{ab}, M^{cd}] = -i(\eta^{ac}M^{bd} + \eta^{bd}M^{ac} - \eta^{ad}M^{bc} - \eta^{bc}M^{ad}) \tag{1.4}$$

where η^{ab} is the diagonal metric tensor such that $\eta^{00} = -\eta^{11} = -\eta^{22} = -\eta^{33} = -\eta^{44} = 1$. The *definition* of the AdS symmetry on quantum level is given by the same equations but $\eta^{44} = 1$.

With such a definition of symmetry on quantum level, dS and AdS symmetries look more natural than Poincare symmetry. In the dS and AdS cases all the ten representation operators of the symmetry algebra are angular momenta while in the Poincare case only six of them are angular momenta and the remaining four operators represent standard energy and momentum. If we define the operators P^μ as $P^\mu = M^{4\mu}/R$ where R is a parameter with the dimension *length* then in the formal limit when $R \to \infty$, $M^{4\mu} \to \infty$ but the quantities P^μ are finite, the relations (1.4) become the relations (1.3). This procedure is called contraction and a general notion of contraction has been proposed in Ref. [21]. In the given case the contraction procedure is the same regardless of whether the relations (1.4) are considered for the dS or AdS symmetry. Note also that the above definitions of the dS and AdS symmetries has nothing to do with dS and AdS spaces and their curvatures.

In view of the above remarks, one might think that the dS analog of the energy operator is M^{40}. However, in dS theory all the operators M^{a0} $(a = 1, 2, 3, 4)$ are on equal footing. This poses a problem whether a parameter describing the evolution defined by the Hamiltonian is a fundamental quantity even on classical level.

In the existing quantum theory, problems with nonphysical notions and infinities arise as a result of describing interactions in terms of local quantum fields. In the present work local quantum fields are not used at all and we apply the notion of symmetry on quantum level only to systems of free particles. One might think that such a consideration can be only of academic interest. Nevertheless, we will see below that there is a class of problems where such a consideration gives a new perspective on fundamental notions of quantum theory. We will consider applications of our approach to the cosmological constant problem, gravity and particle theory.

Finally, let us define the notion of elementary particle. Although theory of elementary particles exists for a rather long period of time, there is no commonly accepted definition of elementary particle in this theory. In the spirit of the above definition of symmetry on quantum level and Wigner's approach to Poincare symmetry [22], a general definition, not depending on the choice of the classical background and on whether we consider a local or nonlocal theory, is that a particle is elementary if the set of its wave functions is the space of an IR of the symmetry algebra in the given theory. In particular, in Poincare invariant theory an elementary particle is described by an IR of the Poincare algebra, in dS or AdS theory it is described by an IR of the dS or AdS algebra, respectively, etc.

1.4 Remarks on the cosmological constant problem

The discovery of the cosmological repulsion (see e.g. Refs. [23, 24]) has ignited a vast discussion on how this phenomenon should be interpreted. The majority of authors treat this phenomenon as an indication that Λ is positive and therefore the space-time background has a positive curvature. According to Refs. [23, 24, 25, 26], the observational data on the value of Λ indicate that it is non-zero and positive with a confidence of 99%. Therefore the possibilities that $\Lambda = 0$ or $\Lambda < 0$ are practically excluded. In the approach discussed in Ref. [27], the "fundamental" quantity Λ is negative while effectively $\Lambda > 0$ only on classical level. In our approach the notion of "fundamental" Λ does not exist since we proceed from the commutation relations (1.4) which do not contain space-time characteristics. We will see below that in our approach Λ arises only in classical approximation. The majority of works dealing with the CC problem proceed from the assumption that G is the fundamental physical quantity, the goal of the theory is to express Λ in terms of G and to explain why Λ is so small.

To consider the CC problem in greater details, we first discuss the following well-known problem: how many independent dimensionful constants are needed for a complete description of nature? A paper [28] represents a trialogue between three well-known scientists: M.J. Duff, L.B. Okun and G. Veneziano (see also Ref. [29] and references therein). The results of their discussions are summarized as follows: *LBO develops the traditional approach with three constants, GV argues in favor of at most two (within superstring theory), while MJD advocates zero.* According to Ref. [30], a possible definition of a fundamental constant might be such that it cannot be calculated in the existing theory. We would like to give arguments in favor of the opinion of the first author in Ref. [28]. One of our goals is to argue that the cosmological and gravitational constants cannot be fundamental physical quantities.

Consider a measurement of a component of angular momentum. The result depends on the system of units. As shown in quantum theory, in units $\hbar/2 = 1$ the result is given by an integer $0, \pm 1, \pm 2, ...$ But we can reverse the order of units and say that in units where the angular momentum is an integer l, its value in $kg \cdot m^2/sec$ is $(1.05457162 \cdot 10^{-34} \cdot l/2) kg \cdot m^2/sec$. Which of those two values has more physical significance? In units where the angular momentum components are integers, the commutation relations between the components are

$$[M_x, M_y] = 2iM_z \quad [M_z, M_x] = 2iM_y \quad [M_y, M_z] = 2iM_x$$

and they do not depend on any parameters. Then the meaning of l is clear: it shows how big the angular momentum is in comparison with the minimum nonzero value 1. At the same time, the measurement of the angular momentum in units $kg \cdot m^2/sec$ reflects only a historic fact that at macroscopic conditions on the Earth in the period between the 18th and 21st centuries people measured the angular momentum in such units.

The fact that quantum theory can be written without the quantity $\hbar$ at all is usually treated as a choice of units where $\hbar = 1/2$ (or $\hbar = 1$). We believe that a better interpretation of this fact is simply that quantum theory tells us that physical results for measurements of the components of angular momentum should be given in integers. Then the question why $\hbar$ is as it is, is not a matter of fundamental physics since the answer is: because we want to measure components of angular momentum in $kg \cdot m^2/sec$.

Our next example is the measurement of velocity v. Let $(E, \mathbf{p})$ be a particle four-momentum defined by its energy and momentum. Then in special relativity the quantity $I_{2P} = E^2 - \mathbf{p}^2c^2$ is an invariant which is denoted as m^2c^4. The reason is that in usual situations $I_{2P} \geq 0$ and m coincides with the standard particle mass. However, if we deal only with four-momenta and don't involve classical space-time then the mathematical structure of Special Relativity does not impose any restrictions on the values of observable quantities E and $\mathbf{p}$; in particular it does not prohibit the case $I_{2P} < 0$. Particles for which this case takes place are called tachyons and their possible existence is widely discussed in the literature. The velocity vector $\mathbf{v}$ is defined as $\mathbf{v} = \mathbf{p}c^2/E$. The fact that any relativistic theory can be written without involving c is usually described as a choice of units where $c = 1$. Then for known particles the quantity $v = |\mathbf{v}|$ can take only values in the range [0,1] while for tachyons it can take values in the range $(1, \infty)$. However, we can again reverse the order of units and say that relativistic theory tells us that for known particles the results for measurements of velocity should be given by values in [0,1] while in general they should be given by values in $[0, \infty)$. Then the question of why c is as it is, is again not a matter of physics since the answer is: because we want to measure velocity in m/sec.

One might pose a question whether or not the values of $\hbar$ and c may change with time. As far as $\hbar$ is concerned, this is a question that if the angular momentum equals one then its value in $kg \cdot m^2/sec$ will always be $1.05457162 \cdot 10^{-34}/2$ or not. It is obvious that this is not a problem of fundamental physics but a problem of definition of the units (kg, m, sec). In other words, this is a problem of metrology and cosmology. At the same time, the value of c will always be the same since the modern *definition* of meter is the length which light passes during $(1/(3 \cdot 10^8))sec$.

It is often stated that the most fundamental constants of nature are $\hbar$, c and G. The units where $\hbar = c = G = 1$ are called Planck units. Another well-known notion is the $c\hbar G$ cube of physical theories. The meaning is that any relativistic theory should contain c, any quantum theory should contain $\hbar$ and any gravitational theory should contain G. However, the above remarks indicates that the meaning should be the opposite. In particular, relativistic theory *should not* contain c and quantum theory *should not contain* $\hbar$. The problem of treating G is one of key problems of this work and will be discussed below.

A standard phrase that relativistic theory becomes non-relativistic one when $c \to \infty$ should be understood such that if relativistic theory is rewritten in conventional (but not physical!) units then c will appear and one can take the limit $c \to \infty$. A more physical description of the transition is that all the velocities in question are much less than unity. We will see in Section 3.6 that those definitions are not equivalent. Analogously, a more physical description of the transition from quantum to classical theory should be that all angular momenta in question are very large rather than $\hbar \to 0$.

Consider now what happens if one assumes that dS symmetry is fundamental. As explained in the preceding section, in our approach dS symmetry has nothing to do with dS space but now we consider standard notion of this symmetry. The dS space is a four-dimensional manifold in the five-dimensional space defined by

$$x_1^2 + x_2^2 + x_3^2 + x_4^2 - x_0^2 = R^2 \tag{1.5}$$

In the formal limit $R \to \infty$ the action of the dS group in a vicinity of the point $(0, 0, 0, 0, x_4 = R)$ becomes the action of the Poincare group on Minkowski space. In the literature, instead of R, the CC $\Lambda = 3/R^2$ is often used. The dS space can be parameterized without using the quantity R at all if instead of x_a $(a = 0, 1, 2, 3, 4)$ we define dimensionless variables $\xi_a = x_a/R$. It is also clear that the elements of the SO(1,4) group do not depend on R since they are products of conventional and hyperbolic rotations. So the dimensionful value of R appears only if one wishes to measure coordinates on the dS space in terms of coordinates of the flat five-dimensional space where the dS space is embedded in. This requirement does not have a fundamental physical meaning. Therefore the value of R defines only a scale factor for measuring coordinates in the dS space. By analogy with c and $\hbar$, the question of why R is as it is, is not a matter of fundamental physics since the answer is: because we want to measure distances in meters. In particular, there is no guaranty that the CC is really a constant, i.e. does not change with time. It is also obvious that if dS symmetry is assumed from the beginning then the value of Λ has no relation to the value of G.

If one assumes that the space-time background is fundamental regardless of whether matter is present or not, then in the spirit of GR it is natural to think that the empty space-time background is flat, i.e. that $\Lambda = 0$ and this was one of the subjects of the well-known debate between Einstein and de Sitter. However, as noted above, it is now accepted that $\Lambda \neq 0$ and, although it is very small, it is positive rather than negative. If we accept parameterization of the dS space as in Eq. (1.5) then the metric tensor on the dS space is

$$g_{\mu\nu} = \eta_{\mu\nu} - x_\mu x_\nu/(R^2 + x_\rho x^\rho) \tag{1.6}$$

where $\mu, \nu, \rho = 0, 1, 2, 3$, $\eta_{\mu\nu}$ is the Minkowski metric tensor, and a summation over repeated indices is assumed. It is easy to calculate the Christoffel symbols in the approximation where all the components of the vector x are much less than R: $\Gamma_{\mu,\nu\rho} = -x_\mu \eta_{\nu\rho}/R^2$. Then a direct calculation shows that in the nonrelativistic approximation the equation of motion for a single particle is

$$\mathbf{a} = \mathbf{r}c^2/R^2 \tag{1.7}$$

where $\mathbf{a}$ and $\mathbf{r}$ are the acceleration and the radius vector of the particle, respectively.

Suppose now that we have a system of two noninteracting particles and $(\mathbf{r}_i, \mathbf{a}_i)$ $(i = 1, 2)$ are their radius vectors and accelerations, respectively. Then Eq. (1.7) is valid for each particle if $(\mathbf{r}, \mathbf{a})$ is replaced by $(\mathbf{r}_i, \mathbf{a}_i)$, respectively. Now if we define the relative radius vector $\mathbf{r} = \mathbf{r}_1 - \mathbf{r}_2$ and the relative acceleration $\mathbf{a} = \mathbf{a}_1 - \mathbf{a}_2$ then they will satisfy the same Eq. (1.7) which shows that the dS antigravity is repulsive. It terms of Λ it reads $\mathbf{a} = \Lambda \mathbf{r}c^2/3$ and therefore in the AdS case we have attraction rather than repulsion.

The fact that even a single particle in the World has a nonzero acceleration might be treated as contradicting the law of inertia but, as already noted, this law has been postulated only for Galilean or Poincare symmetries and we have $\mathbf{a} = 0$ in the limit $R \to \infty$. A more serious problem is that, according to standard experience, any particle moving with acceleration necessarily emits gravitational waves, any charged particle emits electromagnetic waves etc. Does this experience work in the dS world? This problem is intensively discussed in the literature (see e.g. Ref. [31] and references therein). Suppose we accept that, according to GR, the loss of energy in gravitational emission is proportional to the gravitational constant. Then one might say that in the given case it is not legitimate to apply GR since the constant G characterizes interaction between different particles and cannot be used if only one particle exists in the world.

In textbooks on gravity written before 1998 (when the cosmological acceleration was discovered) it is often claimed that Λ is not needed since its presence contradicts the philosophy of GR: matter creates curvature of space-time, so in the absence of matter space-time should be flat (i.e. Minkowski) while empty dS space is not flat. As noted above, such a philosophy has no physical meaning since the notion of empty space-time is unphysical. That's why the discovery of the fact that $\Lambda \neq 0$ has ignited many discussions. The most popular approach is as follows. One can move the term with Λ in Eq. (1.1) from the left-hand side to the right-hand one. Then the term with Λ is treated as the stress-energy tensor of a hidden matter which is called dark energy: $(8\pi G/c^4)T_{\mu\nu}^{DE} = -\Lambda g_{\mu\nu}$. With such an approach one implicitly returns to Einstein's point of view that a curved space-time cannot be empty. In other words, this is an assumption that the Poincare symmetry is fundamental while the dS one is emergent. With the observed value of Λ this dark energy contains approximately 75% of the energy of the World. In this approach G is treated as a fundamental constant and one might try to express Λ in terms of G. The existing quantum theory of gravity cannot perform this calculation unambiguously since the theory contains strong divergences. With a reasonable cutoff parameter, the result for Λ is such that in units where $\hbar = c = 1$, $G\Lambda$ is of the order of unity. This result is expected from dimensionful considerations since in these units, the dimension of G is $length^2$ while the dimension of Λ is $1/length^2$. However, this value of Λ is greater than the observed one by 122 orders of magnitude. In supergravity the disagreement can be reduced but even in best scenarios it exceeds 40 orders of magnitude.This problem is called the CC problem or dark energy problem.

Several authors criticized this approach from the following considerations. GR without the contribution of Λ has been confirmed with a good accuracy in experiments in the Solar System. If Λ is as small as it has been observed then it can have a significant effect only at cosmological distances while for experiments in the Solar System the role of such a small value is negligible. The authors of Ref. [32] titled "Why All These Prejudices Against a Constant?", note that it is not clear why we should think that only a special case $\Lambda = 0$ is allowed. If we accept the theory containing a constant G which cannot be calculated and is taken from the outside then why can't we accept a theory containing two independent constants?

In Secs. 3.6 and 5.1 we show by different methods that, as a consequence of dS symmetry on quantum level defined in the preceding section, the CC problem does not exist and the cosmological acceleration can be easily and naturally explained from first principles of quantum theory.

Concluding this section we note the following. As follows from Eq. (1.7), the quantity R can be extracted from measurements of the relative acceleration in the dS world. However, as follows from this equation, the acceleration is not negligible only if distances between particles are comparable to R. Hence at present a direct measurement of R is impossible and conclusions about its value are made indirectly from the data on high-redshift supernovae by using different cosmological models. Probably the most often used model is the ΛCDM one which is based on six parameters. It assumes that GR is the correct theory of gravity on cosmological scales and uses the FLRW metric (see e.g. Ref. [33]). Then the result of Refs. [25, 26] is that with the accuracy of 5% Λ is such that R is of the order of $10^{26}m$. This value is also obtained in other cosmological models. For example, in the Netchitailo World-Universe model [34] which is based on two parameters, it is adopted that the average density of the world always equals the critical density in GR. Then R is also of the order of $10^{26}m$. On the other hands, in the literature several alternative models are discussed where R considerably differs from $10^{26}m$. It is also important to note that if Λ is treated only as an effective cosmological constant (arising e.g. due to dark energy) then the radius of the world does not define the curvature of the dS space. In summary, in what follows we will treat the fact that $\Lambda > 0$ as a manifestation of dS symmetry on quantum level. On the other hand, the numerical value of R is still an open problem.

1.5 Is the notion of interaction physical?

The fact that problems of quantum theory arise as a result of describing interactions in terms of local quantum fields poses the following dilemma. One can either modify the description of interactions (e.g. by analogy with the string theory where interactions at points are replaced by interactions at strings) or investigate whether the notion of interaction is needed at all. A reader might immediately conclude that the second option fully contradicts the existing knowledge and should be rejected right away. In the present section we discuss a question whether gravity might be not an interaction but simply a kinematical manifestation of dS symmetry on quantum level.

Let us consider an isolated system of two particles and pose a question of whether they interact or not. In theoretical physics there is no unambiguous criterion for answering this question. For example, in classical (i.e. nonquantum) nonrelativistic and relativistic mechanics the criterion is clear and simple: if the relative acceleration of the particles is zero they do not interact, otherwise they interact. However, those theories are based on Galilei and Poincare symmetries, respectively and there is no reason to believe that such symmetries are exact symmetries of nature.

In quantum mechanics the criterion can be as follows. If E is the energy operator of the two-particle system and E_i $(i = 1, 2)$ is the energy operator of particle i then one can formally define the interaction operator U such that

$$E = E_1 + E_2 + U \tag{1.8}$$

Therefore the criterion can be such that the particles do not interact if $U = 0$, i.e. $E = E_1 + E_2$.

In QFT the criterion is also clear and simple: the particles interact if they can exchange by virtual quanta of some fields. For example, the electromagnetic interaction between the particles means that they can exchange by virtual photons, the gravitational interaction - that they can exchange by virtual gravitons etc. In that case U in Eq. (1.8) is an effective operator obtained in the approximation when all degrees of freedom except those corresponding to the given particles can be integrated out.

A problem with approaches based on Eq. (1.8) is that the answer should be given in terms of invariant quantities while energies are reference frame dependent. Therefore one should consider the two-particle mass operator. In standard Poincare invariant theory the free mass operator is given by $M = M_0(\mathbf{q}) = (m_1^2+\mathbf{q}^2)^{1/2}+(m_2^2+\mathbf{q}^2)^{1/2}$ where the m_i are the particle masses and $\mathbf{q}$ is the relative momentum operator. In classical approximation $\mathbf{q}$ becomes the relative momentum and M_0 becomes a function of $\mathbf{q}$ not depending on the relative distance r between the particles. Therefore the relative acceleration is zero and this case can be treated as noninteracting.

Consider now a two-particle system in dS invariant theory. As explained in Sec. 1.3, on quantum level the only consistent definition of dS invariance is that the operators describing the system satisfy the commutation relations of the dS algebra. This definition does not involve GR, QFT, dS space and its geometry (metric, connection etc.). A definition of an elementary particle given in that section is that the particle is described by an IR of the dS algebra (see also Secs. 3.2 and 9.1). Therefore a possible definition of the free two-particle system can be such that the system is described by a representation where not only the energy but all other operators are given by sums of the corresponding single-particle operators. In representation theory such a representation is called the tensor products of IRs.

In other words, we consider only quantum mechanics of two free particles in dS invariant theory. In that case, as shown in Refs. [35, 36, 37] and others (see also Sect. 3.6 of the present work), the two-particle mass operator can be explicitly calculated. It can be written as $M = M_0(\mathbf{q}) + V$ where V is an operator depending not only on $\mathbf{q}$. In classical approximation V becomes a function depending on r. As a consequence, the relative acceleration is not zero and the result for the relative acceleration describes a well-known cosmological repulsion (sometimes called dS anti-gravity). From a formal point of view this result coincides with that obtained in GR on dS space-time (see the preceding section). However, our result has been obtained without involving Riemannian geometry, metric, connection and dS space-time.

One might argue that the above situation contradicts the law of inertia according to which if particles do not interact then their relative acceleration must be zero. However, this law has been postulated in Galilei and Poincare invariant theories and there is no reason to believe that it will be valid for other symmetries. Another argument might be such that dS invariance implicitly implies existence of other particles which interact with the two particles under consideration. Therefore the above situation resembles a case when two particles not interacting with each other are moving with different accelerations in a nonhomogeneous field and therefore their relative acceleration is not zero. This argument has much in common with the discussion of whether the empty space-time background can have a curvature and whether a nonzero curvature implies the existence of dark energy or other fields (see the preceding section). However, as argued in the preceding sections, fundamental quantum theory should not involve the empty space-time background at all. Therefore our result demonstrates that the cosmological constant problem does not exist and the cosmological acceleration can be easily (and naturally) explained without involving dark energy or other fields.

In QFT interactions can be only local and there are no interactions at a distance (sometimes called direct interactions), when particles interact without an intermediate field. In particular, a potential interaction (when the force of the interaction depends only on the distance between the particles) can be only a good approximation in situations when the particle velocities are much less than c. The explanation is such that if the force of the interaction depends only on the distance between the particles and the distance is slightly changed then the particles will feel the change immediately, but this contradicts the statement that no interaction can be transmitted with the speed greater than the speed of light. Although standard QFT is based on Poincare symmetry, physicists typically believe that the notion of interaction adopted in QFT is valid for any symmetry. However, the above discussion shows that the dS antigravity is not caused by exchange of any virtual particles. In particular a question about the speed of propagation of dS antigravity in not physical. In other words, the dS antigravity is an example of a true direct interaction. It is also possible to say that the dS antigravity is not an interaction at all but simply an inherent property of dS invariance.

In quantum theory, dS and AdS symmetries are widely used for investigating QFT in curved space-time background. However, it seems rather paradoxical that such a simple case as a free two-body system in dS invariant theory has not been widely discussed. According to our observations, such a situation is a manifestation of the fact that even physicists working on dS QFT are not familiar with basic facts about IRs of the dS algebra. It is difficult to imagine how standard Poincare invariant quantum theory can be constructed without involving well-known results on IRs of the Poincare algebra. Therefore it is reasonable to think that when Poincare invariance is replaced by dS one, IRs of the Poincare algebra should be replaced by IRs of the dS algebra. However, physicists working on QFT in curved space-time argue that fields are more fundamental than particles and therefore there is no need to involve commutation relations (1.4) and IRs. In other words, they treat dS symmetry on quantum level not such that the relations (1.4) should be valid but such that quantum fields are constructed on dS space (see e.g. Refs. [38, 39]).

Our discussion shows that the notion of interaction depends on symmetry. For example, when we consider a system of two particles which from the point of view of dS symmetry are free (since they are described by a tensor product of IRs), from the point of view of our experience based on Galilei or Poincare symmetries they are not free since their relative acceleration is not zero. This poses a question of whether not only dS antigravity but other interactions are in fact not interactions but effective interactions emerging when a higher symmetry is treated in terms of a lower one.

In particular, is it possible that quantum symmetry is such that on classical level the relative acceleration of two free particles is described by the same expression as that given by the Newton gravitational law and corrections to it? This possibility has been first discussed in Ref. [35]. It is clear that this possibility is not in mainstream according to which gravity is a manifestation of the graviton exchange. We will not discuss whether or not the results on binary pulsars can be treated as a strong indirect indication of the existence of gravitons and why gravitons have not been experimentally detected yet. We believe that until the nature of gravity has been unambiguously understood, different possibilities should be investigated. We believe that a very strong argument in favor of our approach is as follows. In contrast to theories based on Poincare and AdS symmetries, in the dS case the spectrum of the free mass operator is not bounded below by $(m_1 + m_2)$. As a consequence, it is not a problem to indicate states where the mean value of the mass operator has an additional contribution $-Gm_1m_2/r$ with possible corrections. A problem is to understand reasons why macroscopic bodies have such wave functions.

If we accept dS symmetry then the first step is to investigate the structure of dS invariant theory from the point of view of IRs of the dS algebra. This problem is discussed in Refs. [36, 37, 17]. In Ref. [35] we discussed a possibility that gravity is simply a manifestation of the fact that fundamental quantum theory should be based not on complex numbers but on a Galois field with a large characteristic p which is a fundamental constant defining the laws of physics in our World. This approach has been discussed in Refs. [40, 41, 42, 43] and other publications. In Refs. [44, 45] we discussed additional arguments in favor of our hypothesis about gravity. We believe that the results of the present work give strong indications that our hypothesis is correct.

Another arguments that gravity is not an interaction at all follow. The quantity G defines the gravitational force in the Newton law of gravity. Numerous experimental data show that this law works with a very high accuracy. However, this only means that G is a good *phenomenological* parameter. At the level of the Newton law one cannot prove that G is the exact constant which does not change with time, does not depend on masses, distances etc.

General Relativity is a classical (i.e. non-quantum) theory based on the minimum action principle. Here we have two different quantities which have different dimensions: the stress energy tensor of matter and the Ricci tensor describing the curvature of the space-time background. Then the Einstein equations (1.1) derived from the minimum action principle show that G is the coefficient of proportionality between the left-hand and right-hand sides of Eq. (1.1). General Relativity cannot calculate it or give a *theoretical* explanation why this value should be as it is.

A problem arises whether G should be treated as a fundamental or phenomenological constant. By analogy with the treatment of the quantities c and $\hbar$ in the preceding section, one might think that G can be treated analogously and its value is as it is simply because we wish to measure masses in kilograms and distances in meters (in the spirit of Planck units). However, treating G as a fundamental constant can be justified only if there are strong reasons to believe that the Lagrangian of GR is the only possible Lagrangian. Let us consider whether this is the case.

The Lagrangian of GR should be invariant under general coordinate transformations and the simplest way to satisfy this requirement is a choice when it is proportional to the scalar curvature R_c. In this case the Newton gravitational law is recovered in the nonrelativistic approximation and the theory is successful in explaining several well-known phenomena. However, the argument that this choice is simple and agrees with the data, cannot be treated as a fundamental requirement. Another reason for choosing the linear case is that here equations of motions are of the second order while in quadratic, cubic cases etc. they will be of higher orders. However, this reason also cannot be treated as fundamental. It has been argued in the literature that GR is a low energy approximation of a theory where equations of motion contain higher order derivatives. In particular, a rather popular approach is when the Lagrangian contains a function $f(R_c)$ which should be defined from additional considerations. In that case the constant G in the Lagrangian is not the same as the standard gravitational constant. It is believed that the nature of gravity will be understood in the future quantum theory of gravity but efforts to construct this theory has not been successful yet. Therefore the above remarks show that there are no solid reasons to treat G as a fundamental constant.

From the point of view of dS symmetry on quantum level, G cannot be a fundamental constant from the following considerations. The commutation relations (1.4) do not depend on any free parameters. One might say that this is a consequence of the choice of units where $\hbar = c = 1$. However, as noted in the preceding section, any fundamental theory should not involve the quantities $\hbar$ and c. A theory based on the above definition of the dS symmetry on quantum level cannot involve quantities which are dimensionful in units $\hbar = c = 1$. In particular, we inevitably come to conclusion that the gravitational and cosmological constants cannot be fundamental.

By analogy with the above discussion about gravity, one can pose a question of whether the notions of other interactions are fundamental or not. In QFT all interactions (e.g. in QED, electroweak theory and QCD) are introduced according to the same scheme. One writes the Lagrangian as a sum of free and interaction Lagrangians. The latter are proportional to interaction constants which cannot be calculated from the theory and hence can be treated only as phenomenological parameters. It is reasonable to believe that the future fundamental theory will not involve such parameters. For example, one of the ideas of the string theory is that the existing interactions are only manifestations of how higher dimensions are compactified.

1.6 The content of this work

In Chap. 2 we show that in standard nonrelativistic and relativistic quantum theory the position operator is defined inconsistently. As a consequence, in standard quantum theory there exist several paradoxes discussed in Sec. 2.9. We propose a consistent definition of the position operator which resolves the paradoxes and gives a new look at the construction of quantum theory.

In Chap. 3 we construct IRs of the dS algebra following the book by Mensky [46]. This construction makes it possible to show that the well-known cosmological repulsion is simply a kinematical effect in dS quantum mechanics. The derivation involves only standard quantum mechanical notions. It does not require dealing with dS space, metric tensor, connection and other notions of Riemannian geometry. As argued in the preceding sections, fundamental quantum theory should not involve space-time at all. In our approach the cosmological constant problem does not exist and there is no need to involve dark energy or other fields for explaining this problem.

In Chap. 4 we construct IRs of the dS algebra in the basis where all quantum numbers are discrete. In particular, the results of Chap. 2 on the position operator and wave packet spreading are generalized to the dS case. This makes it possible to investigate in Chap. 5 for which two-body wave functions one can get standard Newton's law of gravity and the results which are treated as three classical tests of GR.

In Chap. 6 we argue that fundamental quantum theory should be based on a Galois field rather than complex numbers. In our approach, standard theory is a special case of a quantum theory over a Galois field (GFQT) in a formal limit when the characteristic of the field p becomes infinitely large. We try to make the presentation as self-contained as possible without assuming that the reader is familiar with Galois fields.

In Chap. 7 we construct semiclassical states in GFQT and discuss the problem of calculating the gravitational constant.

In Chap. 8 the AdS symmetry over a Galois field is applied to particle theory. It is shown that in this approach there are no neutral elementary particles in the theory. In particular, even the photon cannot be elementary. The notion of a particle and its antiparticle can be only approximate and such additive quantum numbers as the electric charge and the baryon and lepton quantum numbers can be only approximately conserved.

In Chap. 9 we discuss Dirac singletons in GFQT. Our consideration can be treated as a strong argument in favor of the possibility that only Dirac singletons are true elementary particles.

Finally, Chap. 10 is a discussion.

Chapter 2

A new look at the position operator in quantum theory

2.1 Why do we need position operator in quantum theory?

It has been postulated from the beginning of quantum theory that the coordinate and momentum representations of wave functions are related to each other by a Fourier transform. The historical reason was that in classical electrodynamics the coordinate and wave vector $\mathbf{k}$ representations are related analogously and we postulate that $\mathbf{p} = \hbar\mathbf{k}$ where $\mathbf{p}$ is the particle momentum. Then, although the interpretations of classical fields on one hand and wave functions on the other are fully different, from mathematical point of view classical electrodynamics and quantum mechanics have much in common (and such a situation does not seem to be natural).

One of the examples follows. As explained in textbooks on quantum mechanics (see e.g. Ref. [47]), if the coordinate wave function $\psi(\mathbf{r}, t)$ contains a rapidly oscillating factor $exp[iS(\mathbf{r}, t)/\hbar]$, where $S(\mathbf{r}, t)$ is the classical action as a function of coordinates and time, then in the formal limit $\hbar \to 0$, called semiclassical approximation, the Schrödinger equation becomes the Hamilton-Jacoby equation which shows that quantum mechanical wave packets are moving along classical trajectories. This situation is analogous to the approximation of geometrical optics in classical electrodynamics (see e.g. Ref. [3]) when fields contain a rapidly oscillating factor $exp[i\varphi(\mathbf{r}, t)]$ where the function $\varphi(\mathbf{r}, t)$ is called eikonal. It satisfies the eikonal equation which coincides with the relativistic Hamilton-Jacobi equation for a particle with zero mass. This shows that classical electromagnetic wave packets are moving along classical trajectories for particles with zero mass what is reasonable since it is assumed that such packets consist of photons.

Another example of similarity between classical electrodynamics and quantum mechanics follows. In classical electrodynamics a wave packet moving even in empty space inevitably spreads out and this fact has been known for a long time. For example, as pointed out by Schrödinger (see pp. 41-44 in Ref. [48]), in standard quantum mechanics a packet does not spread out if a particle is moving in a harmonic oscillator potential in contrast to "a wave packet in classical optics, which is dissipated in the course of time". However, as a consequence of the similarity, a free quantum mechanical wave packet inevitably spreads out too. This effect is called wave packet spreading (WPS) and it is described in textbooks and many papers (see e.g. Ref. [49] and references therein). Moreover, as shown in Sec. 2.7, in quantum theory this effect is pronounced even in a much greater extent than in classical electrodynamics.

In particular, the WPS effect has been investigated by de Broglie, Darwin and Schrödinger. The fact that WPS is inevitable has been treated by several authors as unacceptable and as an indication that standard quantum theory should be modified. For example, de Broglie has proposed to describe a free particle not by the Schrödinger equation but by a wavelet which satisfies a nonlinear equation and does not spread out (a detailed description of de Broglie's wavelets can be found e.g. in Ref. [50]). Sapogin writes (see Ref. [51] and references therein) that "Darwin showed that such packet quickly and steadily dissipates and disappears" and proposes an alternative to standard theory which he calls unitary unified quantum field theory.

At the same time, in the literature it has not been explicitly shown that numerical results on WPS are incompatible with experimental data. For example, it is known (see Sec. 2.3) that for macroscopic bodies the effect of WPS is extremely small. Probably it is also believed that in experiments on the Earth with atoms and elementary particles spreading does not have enough time to manifest itself although we have not succeeded in finding an explicit statement on this problem in the literature. Probably for these reasons the majority of physicists do not treat WPS as a drawback of the theory.

However, a natural problem arises what happens to photons which can travel from distant objects to Earth even for billions of years. For example, as shown in Sec. 2.9, in the case when the major part of photons emitted by stars are in wave packet states (what is the most probable scenario) the effect of WPS for photons emitted even by close stars is so strong that we should see not separate stars but rather an almost continuous background from all stars. In addition, data on relic radiation and gamma-ray bursts, signals from radio antennas to planets and space probes, signals from space probes and signals from pulsars show no signs of spreading of photon wave functions. We call those facts the WPS paradoxes. The consideration given in the present chapter shows that the reason of the paradoxes is that standard position operator is not consistently defined. Hence the inconsistent definition of the position operator is not only an academic problem but leads to the above paradoxes.

Usual arguments in favor of choosing the standard position and momentum operators are that these operators have correct properties in semiclassical approximation. For example, in the *method of classical analogy* proposed by Dirac [49] the commutator of operators corresponding to physical quantities is proportional to the classical Poisson bracket of these quantities with the coefficient $i\hbar$. The quantity $\hbar$ has been introduced by Dirac who noted [49] that for agreement with experiment $\hbar$ should be equal to the Planck constant h divided by 2π. Then semiclassical approximation can be treated as a transition from quantum theory to classical one in the formal limit $\hbar \to 0$. However, as noted by Dirac, the method of classical analogy is not universal and in each case when it does not apply special considerations are needed. In addition, the requirement that an operator should have correct properties in semiclassical approximation does not define the operator unambiguously.

One of the arguments in favor of choosing standard position and momentum operators is that the nonrelativistic Schrödinger equation correctly describes the hydrogen energy levels, the Dirac equation correctly describes fine structure corrections to these levels etc. Historically these equations have been first written in coordinate space and in textbooks they are still discussed in this form. However, from the point of view of the present knowledge those equations should be treated as follows.

It is believed that a fundamental theory describing electromagnetic interactions on quantum level is quantum electrodynamics (QED). This theory proceeds from quantizing classical Lagrangian which is only an auxiliary tool for constructing S-matrix. When this construction is accomplished, the results of QED are formulated exclusively in momentum space and the theory does not contain space-time at all. In particular, as follows from the Feynman diagram for the one-photon exchange, in the approximation $(v/c)^2$ the electron in the hydrogen atom can be described in the potential formalism where the potential acts on the wave function in momentum space. So for calculating energy levels one should solve the eigenvalue problem for the Hamiltonian with this potential. This is an integral equation which can be solved by different methods. One of the convenient methods is to apply the Fourier transform and get standard Schrödinger or Dirac equation in coordinate representation with the Coulomb potential. Hence the fact that the results for energy levels are in good agreement with experiment shows only that QED defines the potential correctly and *standard coordinate Schrödinger and Dirac equations are only convenient mathematical ways of solving the eigenvalue problem.* For this problem the physical meaning of the position operator is not important at all. One can consider other transformations of the original integral equation and define other position operators. The fact that for non-standard choices one might obtain something different from the Coulomb potential is not important on quantum level. One might think that on classical level the interaction between two charges can be described by the Coulomb potential but this does not imply that on quantum level the potential in coordinate representation should be necessarily Coulomb.

Let us also note the following. In the literature the statement that the Coulomb law works with a high accuracy is often substantiated from the point of view that predictions of QED have been experimentally confirmed with a high accuracy. However, as follows from the above remarks, the meaning of distance on quantum level is not clear and in QED the law $1/r^2$ can be tested only we assume additionally that the coordinate and momentum representations are related to each other by the Fourier transform. So a conclusion about the validity of the law can be made only on the basis of macroscopic experiments. A conclusion made from the results of classical Cavendish and Maxwell experiments is that if the exponent in Coulomb's law is not 2 but $2 \pm q$ then $q < 1/21600$. The accuracy of those experiments have been considerably improved in the experiment [52] the result of which is $q < 2 \cdot 10^{-9}$. However, the Cavendish-Maxwell experiments and the experiment [52] do not involve pointlike electric charges. Cavendish and Maxwell used a spherical air condenser consisting of two insulated spherical shells while the authors of Ref. [52] developed a technique where the difficulties due to spontaneous ionization and contact potentials were avoided. Therefore the conclusion that $q < 2 \cdot 10^{-9}$ for pointlike electric charges requires additional assumptions.

Another example is as follows. It is said that the spatial distribution of the electric charge inside a system can be extracted from measurements of form-factors in the electron scattering on this system. However, the information about the experiment is again given only in terms of momenta and conclusions about the spatial distribution can be drawn only if we assume additionally how the position operator is expressed in terms of momentum variables. On quantum level the physical meaning of such a spatial distribution is not fundamental.

In quantum theory each elementary particle is described by an irreducible representation (IR) of the symmetry algebra. For example, in Poincare invariant theory the set of momentum operators represents three of ten linearly independent representation operators of the Poincare algebra and hence those operators are consistently defined. On the other hand, among the representation operators there is no position operator. In view of the above discussion, since the *results* of existing fundamental quantum theories describing interactions on quantum level (QED, electroweak theory and QCD) are formulated exclusively in terms of the S-matrix in momentum space without any mentioning of space-time, for investigating such *stationary quantum* problems as calculating energy levels, form-factors etc., the notion of the position operator is not needed.

However, the choice of the position operator is important in nonstationary problems when evolution is described by the time dependent Schrödinger equation (with the nonrelativistic or relativistic Hamiltonian). For any new theory there should exist a correspondence principle that at some conditions the new theory should reproduce results of the old well tested theory with a good accuracy. In particular, quantum theory should reproduce the motion of a particle along the classical trajectory defined by classical equations of motion. Hence *the position operator is needed only in semiclassical approximation* and it should be *defined* from additional considerations.

As noted in Sec. 1.2, in standard approaches to quantum theory the existence of space-time background is assumed from the beginning. Then the position operator for a particle in this background is the operator of multiplication by the particle radius-vector $\mathbf{r}$. As explained in textbooks on quantum mechanics (see e.g. Ref. [47]), the result $-i\hbar\partial/\partial\mathbf{r}$ for the momentum operator can be justified from the requirement that quantum theory should correctly reproduce classical results in semiclassical approximation. However, as noted above, this requirement does not define the operator unambigously.

As noted in Sec. 1.3, the definition of Poincare symmetry on quantum level means that the operators commute according to Eq. (1.3). The fact that an elementary particle in quantum theory is described by an IR of the symmetry algebra can be treated as a definition of the elementary particle (see Sec. 1.3). In Poincare invariant theory the IRs can be implemented in a space of functions $\chi(\mathbf{p})$ such that $\int |\chi(\mathbf{p})|^2 d^3\mathbf{p} < \infty$ (see Sec. 2.4). In this representation the momentum operator $\mathbf{P}$ is defined *unambiguously* and is simply the operator of multiplication by $\mathbf{p}$. A standard *assumption* is that the position operator in this representation is $i\hbar\partial/\partial\mathbf{p}$.

As explained in textbooks on quantum mechanics (see e.g. Ref. [47] and Sec. 2.2), semiclassical approximation cannot be valid in situations when the momentum is rather small. Consider first a one-dimensional case. If the value of the x component of the momentum p_x is rather large, the definition of the coordinate operator $x = i\hbar\partial/\partial p_x$ can be justified but this definition does not have a physical meaning in situations when p_x is small.

Consider now the three-dimensional case. If all the components p_j ($j = 1, 2, 3$) are rather large then there are situations when all the operators $i\hbar\partial/\partial p_j$ are semiclassical. A semiclassical wave function $\chi(\mathbf{p})$ in momentum space should describe a narrow distribution around the mean value $\mathbf{p}_0$. Suppose now that the coordinate axes are chosen such $\mathbf{p}_0$ is directed along the z axis. Then in view of the above remarks the operators $i\hbar\partial/\partial p_j$ cannot be physical for $j = 1, 2$, i.e. in directions perpendicular to the particle momentum. Hence the standard definition of all the components of the position operator can be physical only for special choices of the coordinate axes and there exist choices when the definition is not physical. The situation when a definition of an operator is physical or not depending on the choice of the coordinate axes is not acceptable and hence standard definition of the position operator is not physical.

In the present chapter we propose a consistent definition of the position operator in Poincare invariant theory. As a consequence, in our approach WPS in directions perpendicular to the particle momentum is absent regardless of whether the particle is nonrelativistic or relativistic. Hence the above paradoxes are resolved. Moreover, for an ultrarelativistic particle the effect of WPS is absent at all. In our approach different components of the position operator do not commute with each other and, as a consequence, there is no wave function in coordinate representation.

The chapter is organized as follows. In Secs. 2.2 and 2.4 we discuss the approach to the position operator in standard nonrelativistic and relativistic quantum theory, respectively. An inevitable consequence of this approach is the effect of

WPS of the coordinate wave function which is discussed in Secs. 2.3 and 2.5 for the nonrelativistic and relativistic cases, respectively. In Sec. 2.7 we discuss a relation between the WPS effects for a classical wave packet and for photons comprising this packet. In Sec. 2.8 the problem of WPS in coherent states is discussed. In Sec. 2.9 we show that the WPS effect leads to several paradoxes mentioned above. As discussed in Sec. 2.10, in standard theory it is not possible to avoid those paradoxes. Our approach to a consistent definition of the position operator and its application to WPS are discussed in Secs. 2.11-2.13. Finally, in Sec. 2.14 we discuss implications of the results for entanglement and quantum locality.

2.2 Position operator in nonrelativistic quantum mechanics

In quantum theory, states of a system are represented by elements of a projective Hilbert space. The fact that a Hilbert space H is projective means that if $\psi \in H$ is a state then $const\,\psi$ is the same state. The matter is that not the probability itself but only relative probabilities of different measurement outcomes have a physical meaning. In particular, normalization of states to one is only a matter of convention. This observation will be important in Chaps. 4 and 6 while in this and the next chapters we will always work with states ψ such that $||\psi|| = 1$ where $||...||$ is a norm. It is defined such that if $(...,...)$ is a scalar product in H then $||\psi|| = (\psi,\psi)^{1/2}$.

In quantum theory every physical quantity is described by a self-adjoint operator. Each self-adjoint operator is Hermitian i.e. satisfies the property $(\psi_2, A\psi_1) = (A\psi_2, \psi_1)$ for any states belonging to the domain of A. If A is an operator of some quantity then the mean value of the quantity and its uncertainty in state ψ are given by $\bar{A} = (\psi, A\psi)$ and $\Delta A = ||(A-\bar{A})\psi||$, respectively. The condition that a quantity corresponding to the operator A is semiclassical in state ψ can be defined such that $\Delta A \ll |\bar{A}|$. This implies that the quantity can be semiclassical only if $|\bar{A}|$ is rather large. In particular, if $\bar{A} = 0$ then the quantity cannot be semiclassical.

Let B be an operator corresponding to another physical quantity and $\bar{B}$ and ΔB be the mean value and the uncertainty of this quantity, respectively. We can write $AB = \{A,B\}/2 + [A,B]/2$ where the commutator $[A,B] = AB - BA$ is anti-Hermitian and the anticommutator $\{A,B\} = AB + BA$ is Hermitian. Let $[A,B] = -iC$ and $\bar{C}$ be the mean value of the operator C.

A question arises whether two physical quantities corresponding to the operators A and B can be simultaneously semiclassical in state ψ. Since $||\psi_1||||\psi_2|| \geq |(\psi_1,\psi_2)|$, we have that

$$\Delta A \Delta B \geq \frac{1}{2}|(\psi, (\{A-\bar{A}, B-\bar{B}\} + [A,B])\psi)| \tag{2.1}$$

Since $(\psi, \{A-\bar{A}, B-\bar{B}\}\psi)$ is real and $(\psi, [A,B]\psi)$ is imaginary, we get

$$\Delta A \Delta B \geq \frac{1}{2}|\bar{C}| \tag{2.2}$$

This condition is known as a general uncertainty relation between two quantities. A well-known special case is that if P is the x component of the momentum operator and X is the operator of multiplication by x then $[P,X] = -i\hbar$ and $\Delta p \Delta x \geq \hbar/2$. The states where $\Delta p \Delta = \hbar/2$ are called coherent ones. They are treated such that the momentum and the coordinate are simultaneously semiclassical in a maximal possible way. A well-known example is that if

$$\psi(x) = \frac{1}{a\sqrt{\pi}} exp[\frac{i}{\hbar}p_0 x - \frac{1}{2a^2}(x-x_0)^2]$$

then $\bar{X} = x_0$, $\bar{P} = p_0$, $\Delta x = a/\sqrt{2}$ and $\Delta p = \hbar/(a\sqrt{2})$.

Consider first a one dimensional motion. In standard textbooks on quantum mechanics, the presentation starts with a wave function $\psi(x)$ in coordinate space since it is implicitly assumed that the meaning of space coordinates is known. Then a question arises why $P = -i\hbar d/dx$ should be treated as the momentum operator. The explanation is as follows.

Consider wave functions having the form $\psi(x) = exp(ip_0 x/\hbar)a(x)$ where the amplitude $a(x)$ has a sharp maximum near $x = x_0 \in [x_1, x_2]$ such that $a(x)$ is not small only when $x \in [x_1, x_2]$. Then Δx is of the order of $x_2 - x_1$ and the condition that the coordinate is semiclassical is $\Delta x \ll |x_0|$. Since $-i\hbar d\psi(x)/dx = p_0\psi(x) - i\hbar exp(ip_0 x/\hbar)da(x)/dx$, we see that $\psi(x)$ will be approximately the eigenfunction of $-i\hbar d/dx$ with the eigenvalue p_0 if $|p_0 a(x)| \gg \hbar|da(x)/dx|$. Since $|da(x)/dx|$ is of the order of $|a(x)/\Delta x|$, we have a condition $|p_0\Delta x| \gg \hbar$. Therefore if the momentum operator is $-i\hbar d/dx$, the uncertainty of momentum Δp is of the order of $\hbar/\Delta x$, $|p_0| \gg \Delta p$ and this implies that the momentum is also semiclassical. At the same time, $|p_0\Delta x|/2\pi\hbar$ is approximately the number of oscillations which the exponent makes on the segment $[x_1, x_2]$. Therefore the number of oscillations should be much greater than unity. In particular, semiclassical approximation cannot be valid if Δx is very small, but on the other hand, Δx cannot be very large since it should be much less than x_0. Another justification of the fact that $-i\hbar d/dx$ is the momentum operator is that in the formal limit $\hbar \to 0$ the Schrödinger equation becomes the Hamilton-Jacobi equation. This discussion is similar to a well-known one on the validity of geometrical optics: it is valid when the wave length is much less than characteristic dimensions of the problem.

We conclude that the choice of $-i\hbar d/dx$ as the momentum operator is justified from the requirement that in semiclassical approximation this operator becomes the classical momentum. However, it is obvious that this requirement does not define the operator uniquely: any operator $\tilde{P}$ such that $\tilde{P} - P$ disappears in semiclassical limit, also can be called the momentum operator.

One might say that the choice $P = -i\hbar d/dx$ can also be justified from the following considerations. In nonrelativistic quantum mechanics we assume that the theory should be invariant under the action of the Galilei group, which is a group of transformations of Galilei space-time. The x component of the momentum operator should be the generator corresponding to spatial translations along the x axis and $-i\hbar d/dx$ is precisely the required operator. In this consideration one assumes that space-time has a physical meaning while, as noted in Sect. 1.2, this is not the case.

As noted in Sect. 1.3, one should start not from space-time but from a symmetry algebra. Therefore in nonrelativistic quantum mechanics we should start from the Galilei algebra and consider its IRs. For simplicity we again consider a one dimensional case. Let $P_x = P$ be one of representation operators in an IR of the Galilei algebra. We can implement this IR in a Hilbert space of functions $\chi(p)$ such that $\int_{-\infty}^{\infty} |\chi(p)|^2 dp < \infty$ and P is the operator of multiplication by p, i.e. $P\chi(p) = p\chi(p)$. Then a question arises how the operator of the x coordinate should be defined. In contrast to the momentum operator, the coordinate one is not defined by the representation and so it should be defined from additional assumptions. Probably a future quantum theory of measurements will make it possible to construct operators of physical quantities from the rules how these quantities should be measured. However, at present we can construct necessary operators only from rather intuitive considerations.

By analogy with the above discussion, one can say that semiclassical wave functions should be of the form $\chi(p) = exp(-ix_0 p/\hbar)a(p)$ where the amplitude $a(p)$ has a sharp maximum near $p = p_0 \in [p_1, p_2]$ such that $a(p)$ is not small only when $p \in [p_1, p_2]$. Then Δp is of the order $p_2 - p_1$ and the condition that the momentum is semiclassical is $\Delta p \ll |p_0|$. Since $i\hbar d\chi(p)/dp = x_0\chi(p) + i\hbar exp(-ix_0 p/\hbar)da(p)/dp$, we see that $\chi(p)$ will be approximately the eigenfunction of $i\hbar d/dp$ with the eigenvalue x_0 if $|x_0 a(p)| \gg \hbar |da(p)/dp|$. Since $|da(p)/dp|$ is of the order of $|a(p)/\Delta p|$, we have a condition $|x_0 \Delta p| \gg \hbar$. Therefore if the coordinate operator is $X = i\hbar d/dp$, the uncertainty of coordinate Δx is of the order of $\hbar/\Delta p$, $|x_0| \gg \Delta x$ and this implies that the coordinate defined in such a way is also semiclassical. We can also note that $|x_0 \Delta p|/2\pi\hbar$ is approximately the number of oscillations which the exponent makes on the segment $[p_1, p_2]$ and therefore the number of oscillations should be much greater than unity. It is also clear that semiclassical approximation cannot be valid if Δp is very small, but on the other hand, Δp cannot be very large since it should be much less than p_0.

Although this definition of the coordinate operator has much in common with standard definition of the momentum operators, several questions arise. First of all, by analogy with the discussion about the momentum operator, one can say that the condition that in classical limit the coordinate operator should become the classical coordinate does not define the operator uniquely. One might require that the coordinate operator should correspond to translations in momentum space or be the operator of multiplication by x where the x representation is defined as a

Fourier transform of the p representation but these requirements are not justified. The condition $|x_0| \gg \Delta x$ might seem to be unphysical since x_0 depends on the choice of the origin in the x space while Δx does not depend on this choice. Therefore a conclusion whether the coordinate is semiclassical or not depends on the choice of the reference frame. However, one can notice that not the coordinate itself has a physical meaning but only a relative coordinate between two particles.

Nevertheless, the above definition of the coordinate operator is not fully in line with what we think is a physical coordinate operator. To illustrate this point, consider, for example a measurement of the distance between some particle and the electron in a hydrogen atom. We expect that Δx cannot be less than the Bohr radius. Therefore if x_0 is of the order of the Bohr radius, the coordinate cannot be semiclassical. One might think that the accuracy of the coordinate measurement can be defined as $|\Delta x/x_0|$ and therefore if we succeed in keeping Δx of the order of the Bohr radius when we increase $|x_0|$ then the coordinate will be measured with a better and better accuracy when $|x_0|$ becomes greater. This intuitive understanding might be correct if the distance to the electron is measured in a laboratory where a distance is of the order of centimeters or meters. However, is this intuition correct when we measure distances between macroscopic bodies? In the spirit of GR, the distance between two bodies which are far from each other should be measured by sending a light signal and waiting when it returns back. However, when a reflected signal is obtained, some time has passed and we don't know what happened to the body of interest (e.g. if the body is moving with a high speed, if the World is expanding etc.). For such experiments the logic is opposite to what we have with standard definition of the coordinate operator in quantum mechanics: the accuracy of measurements is better not when the distance is greater but when it is less. One might think that if we consider not very long time intervals then for nonrelativistic particles such a measurement defines the coordinate with a good accuracy. However, it is a problem how to define the distance operator between a macroscopic body and a photon. In view of the remarks in Sect. 1.2 one might think that the photon wave function in coordinate representation might be only a good approximation in semiclassical limit (see also Sec. 2.4).

The above results can be directly generalized to the three-dimensional case. For example, if the coordinate wave function is chosen in the form

$$\psi(\mathbf{r}) = \frac{1}{\pi^{3/4}a^{3/2}} exp[-\frac{(\mathbf{r}-\mathbf{r}_0)^2}{2a^2} + \frac{i}{\hbar}\mathbf{p}_0\mathbf{r}] \tag{2.3}$$

then the momentum wave function is

$$\chi(\mathbf{p}) = \int exp(-\frac{i}{\hbar}\mathbf{p}\mathbf{r})\psi(\mathbf{r})\frac{d^3\mathbf{r}}{(2\pi\hbar)^{3/2}} = \frac{a^{3/2}}{\pi^{3/4}\hbar^{3/2}} exp[-\frac{(\mathbf{p}-\mathbf{p}_0)^2a^2}{2\hbar^2} - \frac{i}{\hbar}(\mathbf{p}-\mathbf{p}_0)\mathbf{r}_0] \tag{2.4}$$

It is easy to verify that

$$||\psi||^2 = \int |\psi(\mathbf{r})|^2 d^3\mathbf{r} = 1, \quad ||\chi||^2 = \int |\chi(\mathbf{p})|^2 d^3\mathbf{p} = 1, \tag{2.5}$$

the uncertainty of each component of the coordinate operator is $a/\sqrt{2}$ and the uncertainty of each component of the momentum operator is $\hbar/(a\sqrt{2})$. Hence one might think that Eqs. (2.3) and (2.4) describe a state which is semiclassical in a maximal possible extent.

Let us make the following remark about semiclassical vector quantities. We defined a quantity as semiclassical if its uncertainty is much less than its mean value. In particular, as noted above, a quantity cannot be semiclassical if its mean value is small. In the case of vector quantities we have sets of three physical quantities. Some of them can be small and for them it is meaningless to discuss whether they are semiclassical or not. We say that a vector quantity is semiclassical if all its components which are not small are semiclassical and there should be at least one semiclassical component.

For example, if the mean value of the momentum $\mathbf{p}_0$ is directed along the z axes then the xy components of the momentum are not semiclassical but the three-dimensional vector quantity $\mathbf{p}$ can be semiclassical if $\mathbf{p}_0$ is rather large. However, in that case the definitions of the x and y components of the position operator as $x = i\hbar\partial/\partial p_x$ and $y = i\hbar\partial/\partial p_y$ become inconsistent. The situation when the validity of an operator depends on the choice of directions of the coordinate axes is not acceptable and hence the above definition of the position operator is at least problematic. Moreover, as already mentioned, it will be shown in Sec. 2.9 that the standard choice of the position operator leads to the WPS paradoxes.

Let us note that semiclassical states can be constructed not only in momentum or coordinate representations. For example, instead of momentum wave functions $\chi(\mathbf{p})$ one can work in the representation where the quantum numbers (p, l, μ) in wave functions $\chi(p, l, \mu)$ mean the magnitude of the momentum p, the orbital quantum number l (such that a state is the eigenstate of the orbital momentum squared $\mathbf{L}^2$ with the eigenvalue $l(l+1)$) and the magnetic quantum number μ (such that a state is the eigenvector or L_z with the eigenvalue μ). A state described by a $\chi(p, l, \mu)$ will be semiclassical with respect to those quantum numbers if $\chi(p, l, \mu)$ has a sharp maximum at $p = p_0$, $l = l_0$, $\mu = \mu_0$ and the widths of the maxima in p, l and μ are much less than p_0, l_0 and μ_0, respectively. However, by analogy with the above discussion, those widths cannot be arbitrarily small if one wishes to have other semiclassical variables (e.g. the coordinates). Examples of such situations will be discussed in Sec. 2.12.

2.3 Wave packet spreading in nonrelativistic quantum mechanics

As noted in Sec. 1.2, in quantum theory there is no operator having the meaning of the time operator and it is usually assumed that time is a classical parameter such that the dependence of the wave function on time is defined by the Hamiltonian according to the Schrödinger equation. As discussed in Sec. 1.2, this treatment of time encounters several problems. However, in this chapter we consider the WPS paradoxes assuming that the standard treatment of time is valid for describing photons and other elementary particles.

In nonrelativistic quantum mechanics the Hamiltonian of a free particle with the mass m is $H = \mathbf{p}^2/2m$ and hence, as follows from Eq. (2.4), in the model discussed above the dependence of the momentum wave function on t is given by

$$\chi(\mathbf{p},t) = \frac{a^{3/2}}{\pi^{3/4}\hbar^{3/2}} exp[-\frac{(\mathbf{p}-\mathbf{p}_0)^2a^2}{2\hbar^2} - \frac{i}{\hbar}(\mathbf{p}-\mathbf{p}_0)\mathbf{r}_0 - \frac{i\mathbf{p}^2t}{2m\hbar}] \tag{2.6}$$

It is easy to verify that for this state the mean value of the operator $\mathbf{p}$ and the uncertainty of each momentum component are the same as for the state $\chi(\mathbf{p})$, i.e. those quantities do not change with time.

Consider now the dependence of the coordinate wave function on t. This dependence can be calculated by using Eq. (2.6) and the fact that

$$\psi(\mathbf{r},t) = \int exp(\frac{i}{\hbar}\mathbf{pr})\chi(\mathbf{p},t)\frac{d^3\mathbf{p}}{(2\pi\hbar)^{3/2}} \tag{2.7}$$

The result of a direct calculation is

$$\psi(\mathbf{r},t) = \frac{1}{\pi^{3/4}a^{3/2}}(1+\frac{i\hbar t}{ma^2})^{-3/2} exp[-\frac{(\mathbf{r}-\mathbf{r}_0-\mathbf{v}_0t)^2}{2a^2(1+\frac{\hbar^2t^2}{m^2a^4})}(1-\frac{i\hbar t}{ma^2}) + \frac{i}{\hbar}\mathbf{p}_0\mathbf{r} - \frac{i\mathbf{p}_0^2t}{2m\hbar}] \tag{2.8}$$

where $\mathbf{v}_0 = \mathbf{p}_0/m$ is the classical velocity. This result shows that the semiclassical wave packet is moving along the classical trajectory $\mathbf{r}(t) = \mathbf{r}_0 + \mathbf{v}_0t$. At the same time, it is now obvious that the uncertainty of each coordinate depends on time as

$$\Delta x_j(t) = \Delta x_j(0)(1+\hbar^2t^2/m^2a^4)^{1/2}, \quad (j = 1,2,3) \tag{2.9}$$

where $\Delta x_j(0) = a/\sqrt{2}$, i.e. the width of the wave packet in coordinate representation is increasing. This fact, known as the wave-packet spreading (WPS), is described in many textbooks and papers (see e.g. the textbooks [49] and references therein). It shows that if a state was semiclassical in the maximal extent at $t = 0$, it will not have this property at $t > 0$ and the accuracy of semiclassical approximation will decrease with the increase of t. The characteristic time of spreading can be defined as $t_* = ma^2/\hbar$. For macroscopic bodies this is an extremely large quantity and hence in macroscopic physics the effect of WPS can be neglected. In the formal limit $\hbar \to 0$, t_* becomes infinite, i.e. spreading does not take place. This shows that WPS is a pure quantum phenomenon. For the first time the result (2.8) has been obtained by Darwin in Ref. [53].

One might pose a problem whether the WPS effect is specific only for Gaussian wave functions. One might expect that this effect will take place in general situations since each component of standard position operator $i\hbar\partial/\partial\mathbf{p}$ does not commute with the Hamiltonian and so the distribution of the corresponding physical quantity will be time dependent. A good example showing inevitability of WPS is as follows. If at $t = 0$ the coordinate wave function is $\psi_0(\mathbf{r})$ then, as follows from Eqs. (2.4) and (2.7),

$$\psi(\mathbf{r},t) = \int exp\{\frac{i}{\hbar}[\mathbf{p}(\mathbf{r}-\mathbf{r}') - \frac{\mathbf{p}^2 t}{2m}]\}\psi_0(\mathbf{r}')\frac{d^3\mathbf{r}'d^3\mathbf{p}}{(2\pi\hbar)^3} \tag{2.10}$$

As follows from this expression, if $\psi_0(\mathbf{r}) \neq 0$ only if $\mathbf{r}$ belongs to a finite vicinity of some vector $\mathbf{r}_0$ then at any $t > 0$ the carrier of $\psi(\mathbf{r},t)$ belongs to the whole three-dimensional space, i.e. the wave function spreads out with an infinite speed. One might think that in nonrelativistic theory this is not unacceptable since this theory can be treated as a formal limit $c \to \infty$ of relativistic theory.

As shown in Ref. [54] titled "Nonspreading wave packets", for a one-dimensional wave function in the form of an Airy function, spreading does not take place and the maximum of the quantity $|\psi(x)|^2$ propagates with constant acceleration even in the absence of external forces. Those properties of Airy packets have been observed in optical experiments [55]. However, since such a wave function is not normalizable, we believe that the term "wave packet" in the given situation might be misleading since the mean values and uncertainties of the coordinate and momentum cannot be calculated in a standard way. Such a wave function can be constructed only in a limited region of space. As explained in Ref. [54], this wave function describes not a particle but rather families of particle orbits. As shown in Ref. [54], one can construct a normalized state which is a superposition of Airy functions with Gaussian coefficients and "eventually the spreading due to the Gaussian cutoff takes over". This is an additional argument that the effect of WPS is an inevitable consequence of standard quantum theory.

Since quantum theory is invariant under time reversal, one might ask the following question: is it possible that the width of the wave packet in coordinate representation is decreasing with time? From the formal point of view, the answer is "yes". Indeed, the solution given by Eq. (2.8) is valid not only when $t \geq 0$ but when $t < 0$ as well. Then, as follows from Eq. (2.9), the uncertainty of each coordinate is decreasing when t changes from some negative value to zero. However, eventually the value of t will become positive and the quantities $\Delta x_j(t)$ will grow to infinity. In this chapter we consider situations when a photon is created on atomic level and hence one might expect that its initial coordinate uncertainties are not large. However, when the photon travels a long distance to the Earth, those uncertainties become much greater, i.e. the term WPS reflects the physics adequately.

2.4 Position operator in relativistic quantum mechanics

The problem of the position operator in relativistic quantum theory has been discussed in a wide literature and different authors have different opinions on this problem. In particular, some authors state that in relativistic quantum theory no position operator exists. As already noted, the results of fundamental quantum theories are formulated only in terms of the S-matrix in momentum space without any mentioning of space-time. This is in the spirit of the Heisenberg S-matrix program that in relativistic quantum theory it is possible to describe only transitions of states from the infinite past when $t \to -\infty$ to the distant future when $t \to +\infty$. On the other hand, since quantum theory is treated as a theory more general than classical one, it is not possible to fully avoid space and time in quantum theory. For example, quantum theory should explain how photons from distant objects travel to Earth and even how macroscopic bodies are moving along classical trajectories. Hence we can conclude that: a) in quantum theory (nonrelativistic and relativistic) we must have a position operator and b) this operator has a physical meaning only in semiclassical approximation.

There exists a wide literature describing how IRs of the Poincare algebra can be constructed. In particular, an IR for a spinless particle can be implemented in a space of functions $\xi(\mathbf{p})$ satisfying the condition

$$\int |\xi(\mathbf{p})|^2 d\rho(\mathbf{p}) < \infty, \quad d\rho(\mathbf{p}) = \frac{d^3\mathbf{p}}{\epsilon(\mathbf{p})} \tag{2.11}$$

where $\epsilon(\mathbf{p}) = (m^2 + \mathbf{p}^2)^{1/2}$ is the energy of the particle with the mass m. The convenience of the above requirement is that the volume element $d\rho(\mathbf{p})$ is Lorentz invariant. In that case it can be easily shown by direct calculations (see e.g. Ref. [56]) that the representation operators have the form

$$\mathbf{L} = -i\mathbf{p} \times \frac{\partial}{\partial \mathbf{p}}, \quad \mathbf{N} = -i\epsilon(\mathbf{p})\frac{\partial}{\partial \mathbf{p}}, \quad \mathbf{P} = \mathbf{p}, \quad E = \epsilon(\mathbf{p}) \tag{2.12}$$

where $\mathbf{L}$ is the orbital angular momentum operator, $\mathbf{N}$ is the Lorentz boost operator, $\mathbf{P}$ is the momentum operator, E is the energy operator and these operators are expressed in terms of the operators in Eq. (1.3) as

$$\mathbf{L} = (M^{23}, M^{31}, M^{12}),\ \mathbf{N} = (M^{10}, M^{20}, M^{30}),\ \mathbf{P} = (P^1, P^2, P^3),\ E = P^0$$

For particles with spin these results are modified as follows. For a massive particle with spin s the functions $\xi(\mathbf{p})$ also depend on spin projections which can take $2s+1$ values $-s, -s+1, ...s$. If $\mathbf{s}$ is the spin operator then the total angular momentum has an additional term $\mathbf{s}$ and the Lorentz boost operator has an additional term $(\mathbf{s}\times\mathbf{p})/(\epsilon(\mathbf{p})+m)$ (see e.g. Eq. (2.5) in Ref. [56]). Hence corrections of the spin terms to the quantum numbers describing the angular momentum and the Lorentz boost do not exceed s. We assume as usual that in semiclassical approximation the quantum numbers characterizing the angular momentum and the Lorentz boost are much greater than unity and hence in this approximation spin effects can be neglected. For a massless particle with the spin s the spin projections can take only values $-s$ and s and those quantum numbers have the meaning of helicity. In this case the results for the representation operators can be obtained by taking the limit $m \to 0$ if the operators are written in the light front variables (see e.g. Eq. (25) in Ref. [17]). As a consequence, in semiclassical approximation the spin corrections in the massless case can be neglected as well. Hence for investigating the position operator we will neglect spin effects and will not explicitly write the dependence of wave functions on spin projections.

In the above IRs the representation operators are Hermitian as it should be for operators corresponding to physical quantities. In standard theory (over complex numbers) such IRs of the Lie algebra can be extended to unitary IRs of the Poincare group. In the literature elementary particles are described not only by such IRs but also by local fields and, as noted in Sec. 1.2, their physical meaning is problematic. Below we discuss the both approaches but first we consider the case of unitaty IRs.

As follows from Eq. (1.3), the operator $I_2 = E^2 - \mathbf{P}^2$ is the Casimir operator of the second order, i.e. it is a bilinear combination of representation operators commuting with all the operators of the algebra. As follows from the known Schur lemma, all states belonging to an IR are the eigenvectors of I_2 with the same eigenvalue m^2. Note that Eq. (2.12) contains only m^2 but not m. The choice of the energy sign is only a matter of convention but not a matter of principle. Indeed, the energy can be measured only if the momentum $\mathbf{p}$ is measured and then it is only a matter of convention what sign of the square root should be chosen. However, it is important that the sign should be the same for all particles. For example, if we consider a system of two particles with the same values of m^2 and the opposite momenta $\mathbf{p}_1$ and $\mathbf{p}_2$ such that $\mathbf{p}_1 + \mathbf{p}_2 = 0$, we cannot define the energies of the particles as $\epsilon(\mathbf{p}_1)$ and $-\epsilon(\mathbf{p}_2)$, respectively, since in that case the total four-momentum of the two-particle system will be zero what contradicts experiment.

The notation $I_2 = m^2$ is justified by the fact that for all known particles I_2 is greater or equal than zero. Then the mass m is *defined* as the square root of m^2 and the sign of m is only a matter of convention. The usual convention is that $m \geq 0$. However, from mathematical point of view, IRs with $I_2 < 0$ are not prohibited. If the velocity operator $\mathbf{v}$ is *defined* as $\mathbf{v} = \mathbf{P}/E$ then for known particles $|\mathbf{v}| \leq 1$, i.e. $|\mathbf{v}| \leq c$ in standard units. However, for IRs with $I_2 < 0$, $|\mathbf{v}| > c$ and, at least from the point of view of mathematical construction of IRs, this case is not prohibited. The hypothetical particles with such properties are called tachyons and their possible existence is widely discussed in the literature. If the tachyon mass m is also defined as the square root of m^2 then this quantity will be imaginary. However, this does not mean than the corresponding IRs are unphysical since all the operators of the

Poincare group Lie algebra depend only on m^2.

As follows from Eqs. (2.11) and (2.12), in the nonrelativistic approximation $d\rho(\mathbf{p}) = d^3\mathbf{p}/m$ and $\mathbf{N} = -im\partial/\partial\mathbf{p}$. Therefore in this approximation $\mathbf{N}$ is proportional to *standard* position operator and one can say that the position operator is in fact present in the description of the IR.

In relativistic case the operator $i\partial/\partial\mathbf{p}$ is not selfadjoint since $d\rho(\mathbf{p})$ is not proportional to $d^3\mathbf{p}$. However, one can perform a unitary transformation $\xi(\mathbf{p}) \to \chi(\mathbf{p}) = \xi(\mathbf{p})/\epsilon(\mathbf{p})^{1/2}$ such that the Hilbert space becomes the space of functions $\chi(\mathbf{p})$ satisfying the condition $\int |\chi(\mathbf{p})|^2 d^3\mathbf{p} < \infty$. It is easy to verify that in this implementation of the IR the operators $(\mathbf{L}, \mathbf{P}, E)$ will have the same form as in Eq. (2.12) but the expression for $\mathbf{N}$ will be

$$\mathbf{N} = -i\epsilon(\mathbf{p})^{1/2}\frac{\partial}{\partial\mathbf{p}}\epsilon(\mathbf{p})^{1/2} \tag{2.13}$$

In this case one can *define* $i\hbar\partial/\partial\mathbf{p}$ as a position operator but now we do not have a situation when the position operator is present among the other representation operators.

A problem of the definition of the position operator in relativistic quantum theory has been discussed since the beginning of the 1930s and it has been noted that when quantum theory is combined with relativity the existence of the position operator with correct physical properties becomes a problem. The above definition has been proposed by Newton and Wigner in Ref. [20]. They worked in the approach when elementary particles are described by local fields rather than unitary IRs. The Fourier transform of such fields describes states where the energy can be positive and negative and this is interpreted such that local quantum fields describe a particle and its antiparticle simultaneously. Newton and Wigner first discuss the spinless case and consider only states on the upper Lorentz hyperboloid where the energy is positive. For such states the representation operators act in the same way as in the case of spinless unitary IRs. With this definition the coordinate wave function $\psi(\mathbf{r})$ can be again defined by Eq. (2.3) and a question arises whether such a position operator has all the required properties.

For example, in the introductory section of the textbook [16] the following arguments are given in favor of the statement that in relativistic quantum theory it is not possible to define a physical position operator. Suppose that we measure coordinates of an electron with the mass m. When the uncertainty of coordinates is of the order of $\hbar/mc$, the uncertainty of momenta is of the order of mc, the uncertainty of energy is of the order of mc^2 and hence creation of electron-positron pairs is allowed. As a consequence, it is not possible to localize the electron with the accuracy better than its Compton wave length $\hbar/mc$. Hence, for a particle with a nonzero mass exact measurement is possible only either in the nonrelativistic limit (when $c \to \infty$) or classical limit (when $\hbar \to 0$). In the case of the photon, as noted by Pauli (see p. 191 of Ref. [7]), the coordinate cannot be measured with the accuracy better than

$\hbar/p$ where p is the magnitude of the photon momentum. The quantity $\lambda = 2\pi\hbar/p$ is called the photon wave length (see Sec. 1.2). Since $\lambda \to 0$ in the formal limit $\hbar \to 0$, Pauli concludes that "Only within the confines of the classical ray concept does the position of the photon have a physical significance".

Another argument that the Newton-Wigner position operator does not have all the required properties follows. A relativistic analog of Eq. (2.10) is

$$\psi(\mathbf{r},t) = \int exp\{\frac{i}{\hbar}[\mathbf{p}(\mathbf{r}-\mathbf{r}') - \epsilon(\mathbf{p})t]\}\psi_0(\mathbf{r}')\frac{d^3\mathbf{r}'d^3\mathbf{p}}{(2\pi\hbar)^3} \tag{2.14}$$

As a consequence, the Newton-Wigner position operator has the "tail property": if $\psi_0(\mathbf{r}) \neq 0$ only if $\mathbf{r}$ belongs to a finite vicinity of some vector $\mathbf{r}_0$ then at any $t > 0$ the function $\psi(\mathbf{r},t)$ has a tail belonging to the whole three-dimensional space, i.e. the wave function spreads out with an infinite speed. Hence at any $t > 0$ the particle can be detected at any point of the space and this contradicts the requirement that no information should be transmitted with the speed greater than c.

The tail property of the Newton-Wigner position operator has been known for a long time (see e.g. Ref. [57] and references therein). It is characterized as non-locality leading to the action at a distance. Hegerfeldt argues [57] that this property is rather general because it can be proved assuming that energy is positive and without assuming a specific choice of the position operator. The Hegerfeldt theorem [57] is based on the assumption that there exists an operator $N(V)$ whose expectation defines the probability to find a particle inside the volume V. However, the meaning of time on quantum level is not clear and for the position operator proposed in the present paper such a probability does not exist because there is no wave function in coordinate representation (see Sec. 2.11 and the discussion in Sec. 2.14).

One might say that the requirement that no signal can be transmitted with the speed greater than c has been obtained in Special Relativity which is a classical (i.e. nonquantum) theory operating only with classical space-time coordinates. For example, in classical theory the velocity of a particle is defined as $\mathbf{v} = d\mathbf{r}/dt$ but, as noted above, the velocity *should be defined* as $\mathbf{v} = \mathbf{p}/E$ (i.e. without mentioning space-time) and then on classical level it can be shown that $\mathbf{v} = d\mathbf{r}/dt$. In QFT local quantum fields separated by space-like intervals commute or anticommute (depending on whether the spin is integer or half-integer) and this is treated as a requirement of causality and that no signal can be transmitted with the speed greater than c. However, as noted above, the physical meaning of space-time coordinates on quantum level is not clear. Hence from the point of view of quantum theory the existence of tachyons is not prohibited. Note also that when two electrically charged particles exchange by a virtual photon, a typical situation is that the four-momentum of the photon is space-like, i.e. the photon is the tachyon. We conclude that although in relativistic theory such a behavior might seem undesirable, there is no proof that it must be excluded. Also, as argued by Griffiths (see Ref. [58] and references therein), with a consistent interpretation of quantum theory there are no nonlocality and superluminal interactions. In Sec. 2.14 we argue that the position operator proposed in the present paper sheds a new light on this problem.

An example with the 21cm transition line between the hyperfine energy levels of the hydrogen atom mentioned in Sec. 1.2 describes a pure quantum phenomenon while, as noted above, a position operator is needed only in semiclassical approximation.

For particles with nonzero spin, the number of states in local fields is typically by a factor of two greater than in the case of unitary IRs (since local fields describe a particle and its antiparticle simultaneously) but those components are not independent since local fields satisfy a covariant equation (Klein-Gordon, Dirac etc.). In Ref. [20] Newton and Wigner construct a position operator in the massive case but say that in the massless one they have succeeded in constructing such an operator only for Klein-Gordon and Dirac particles while in the case of the photon the position operator does not exist. On the other hand, as noted above, in the case of unitary IRs different spin components are independent and in semiclassical approximation spin effects are not important. So in this approach one might adopt the Newton-Wigner position operator for particles with any spin and any mass.

In view of the WPS paradoxes, we consider the photon case in greater details. In textbooks on QED (see e.g. Ref. [12]) it is stated that in this theory there is no way to define a coordinate photon wave function and the arguments are as follows. The electric and magnetic fields of the photon in coordinate representation are proportional to the Fourier transforms of $|\mathbf{p}|^{1/2}\chi(\mathbf{p})$, rather than $\chi(\mathbf{p})$. As a consequence, the quantities $\mathbf{E}(\mathbf{r})$ and $\mathbf{B}(\mathbf{r})$ are defined not by $\psi(\mathbf{r})$ but by integrals of $\psi(\mathbf{r})$ over a region of the order of the wave length. However, this argument also does not exclude the possibility that $\psi(\mathbf{r})$ can have a physical meaning in semiclassical approximation since, as noted in Sec. 1.2, the notions of the electric and magnetic fields of the single photon do not have a physical meaning. In addition, since $\lambda \to 0$ in the formal limit $\hbar \to 0$, one should not expect that any position operator in semiclassical approximation can describe coordinates with the accuracy better than the wave length.

A detailed discussion of the photon position operator can be found in papers by Margaret Hawton and references therein (see e.g. Ref. [59]). In this approach the photon is described by a local field and the momentum and coordinate representations are related to each other by standard Fourier transform. The author of Ref. [59] discusses generalizations of the photon position operator proposed by Pryce [60]. However, the Pryce operator and its generalizations discussed in Ref. [59] differ from the Newton-Wigner operator only by terms of the order of the wave length. Hence in semiclassical approximation all those operators are equivalent.

The above discussion shows that on quantum level the physical meaning of the coordinate is not clear but in view of a) and b) (see the beginning of this section) one can conclude that in semiclassical approximation all the existing proposals for the position operator are equivalent to the Newton-Wigner operator $i\hbar\partial/\partial\mathbf{p}$. An additional argument in favor of this operator is that the relativistic nature of the photon might be somehow manifested in the longitudinal direction while in transverse directions the behavior of the wave function should be similar to that in standard nonrelativistic quantum mechanics. Another argument is that the photon wave function in coordinate representation constructed by using this operator satisfies the wave equation in agreement with classical electrodynamics (see Sec. 2.6).

In addition, if we consider a motion of a free particle, it is not important in what interactions this particle participates and, as explained above, if the particle is described by its IR in semiclassical approximation then the particle spin is not important. Hence the effect of WPS for an ultrarelativistic particle does not depend on the nature of the particle, i.e. on whether the particle is the photon, the proton, the electron etc.

For all the reasons described above and in view of a) and b), in the next section we consider what happens if the space-time evolution of relativistic wave packets is described by using the Newton-Wigner position operator.

2.5 Wave packet spreading in relativistic quantum mechanics

Consider first a construction of the wave packet for a particle with nonzero mass. A possible way of the construction follows. We first consider the particle in its rest system, i.e. in the reference frame where the mean value of the particle momentum is zero. The wave function $\chi_0(\mathbf{p})$ in this case can be taken as in Eq. (2.4) with $\mathbf{p}_0 = 0$. As noted in Sec. 2.2, such a state cannot be semiclassical. However, it is possible to obtain a semiclassical state by applying a Lorentz transformation to $\chi_0(\mathbf{p})$. One can show (see e.g. Eq. (2.4) in Ref. [56]) that when the IR for a spinless particle is extended to the unitary representation of the Poincare group then the operator $U(g)$ corresponding to a Lorentz transformation g is

$$U(g)\chi_0(\mathbf{p}) = [\frac{\epsilon(\mathbf{p}')}{\epsilon(\mathbf{p})}]^{1/2}\chi_0(\mathbf{p}') \tag{2.15}$$

where $\mathbf{p}'$ is the momentum obtained from $\mathbf{p}$ by the Lorentz transformation g^{-1}. If g is the Lorentz boost along the z axis with the velocity v then

$$\mathbf{p}'_\perp = \mathbf{p}_\perp, \quad p'_z = \frac{p_z - v\epsilon(\mathbf{p})}{(1-v^2)^{1/2}} \tag{2.16}$$

where we use the subscript $\perp$ to denote projections of vectors onto the xy plane.

As follows from this expression, $exp(-\mathbf{p}'^2a^2/2\hbar^2)$ as a function of $\mathbf{p}$ has the maximum at $\mathbf{p}_\perp = 0$, $p_z = p_{z0} = v[(m^2+\mathbf{p}_\perp^2)/(1-v^2)]^{1/2}$ and near the maximum

$$exp(-\frac{a^2\mathbf{p}'^2}{2\hbar^2}) \approx exp\{-\frac{1}{2\hbar^2}[a^2\mathbf{p}_\perp^2 + b^2(p_z-p_{z0})^2]\}$$

where $b = a(1-v^2)^{1/2}$ what represents the effect of the Lorentz contraction. If $mv \gg \hbar/a$ (in units where $c = 1$) then $m \gg |\mathbf{p}_\perp|$ and $p_{z0} \approx mv/(1-v^2)^{1/2}$. In this case the transformed state is semiclassical and the mean value of the momentum is exactly the classical (i.e. nonquantum) value of the momentum of a particle with mass m moving along the z axis with the velocity v. However, in the opposite case when $m \ll \hbar/a$ the transformed state is not semiclassical since the uncertainty of p_z is of the same order as the mean value of p_z.

If the photon mass is exactly zero then the photon cannot have the rest state. However, even if the photon mass is not exactly zero, it is so small that the relation $m \ll \hbar/a$ is certainly satisfied for any realistic value of a. Hence a semiclassical state for the photon or a particle with a very small mass cannot be obtained by applying the Lorentz transformation to $\chi_0(\mathbf{p})$ and considering the case when v is very close to unity. In this case we will describe a semiclassical state by a wave function which is a generalization of the function (2.4):

$$\chi(\mathbf{p},0) = \frac{ab^{1/2}}{\pi^{3/4}\hbar^{3/2}} exp[-\frac{\mathbf{p}_\perp^2 a^2}{2\hbar^2} - \frac{(p_z-p_0)^2 b^2}{2\hbar^2} - \frac{i}{\hbar}\mathbf{p}_\perp \mathbf{r}_{0\perp} - \frac{i}{\hbar}(p_z-p_0)z_0] \qquad (2.17)$$

Here we assume that the vector $\mathbf{p}_0$ is directed along the z axis and its z component is p_0. In the general case the parameters a and b defining the momentum distributions in the transverse and longitudinal directions, respectively, can be different. In that case the uncertainty of each transverse component of momentum is $\hbar/(a\sqrt{2})$ while the uncertainty of the z component of momentum is $\hbar/(b\sqrt{2})$. In view of the above discussion one might think that, as a consequence of the Lorentz contraction, the parameter b should be very small. However, the above discussion shows that the notion of the Lorentz contraction has a physical meaning only if $m \gg \hbar/a$ while for the photon the opposite relation takes place. We will see below that in typical situations the quantity b is large and much greater than a.

As noted in Sec. 2.3, in this chapter we assume that in some situations time is a good approximate parameter describing evolution. Hence in the relativistic case evolution is described by the Schrödinger equation with the relativistic Hamiltonian. Then the dependence of the momentum wave function (2.17) on t is given by

$$\chi(\mathbf{p},t) = exp(-\frac{i}{\hbar}pct)\chi(\mathbf{p},0) \qquad (2.18)$$

where $p = |\mathbf{p}|$ and we assume that the particle is ultrarelativistic, i.e. $p \gg m$. Since at different moments of time the wave functions in momentum space differ each other only by a phase factor, the mean value and uncertainty of each momentum component do not depend on time. In other words, there is no WPS for the wave function in momentum space. As noted in Sec. 2.3, the same is true in the nonrelativistic case.

In view of the above discussion, the function $\psi(\mathbf{r},t)$ can be again defined by Eq. (2.7) where now $\chi(\mathbf{p},t)$ is defined by Eq. (2.18). If the variable p_z in the integrand is replaced by $p_0 + p_z$ then as follows from Eqs. (2.7,2.17,2.18)

$$\psi(\mathbf{r},t) = \frac{ab^{1/2}exp(i\mathbf{p}_0\mathbf{r}/\hbar)}{\pi^{3/4}\hbar^{3/2}(2\pi\hbar)^{3/2}}\int exp\{-\frac{\mathbf{p}_\perp^2 a^2}{2\hbar^2} - \frac{p_z^2 b^2}{2\hbar^2} + \frac{i}{\hbar}\mathbf{p}(\mathbf{r}-\mathbf{r}_0)$$
$$-\frac{ict}{\hbar}[(p_z+p_0)^2 + \mathbf{p}_\perp^2]^{1/2}\}d^3\mathbf{p} \tag{2.19}$$

We now take into account the fact that in semiclassical approximation the quantity p_0 should be much greater than the uncertainties of the momentum in the longitudinal and transversal directions, i.e. $p_0 \gg p_z$ and $p_0 \gg |\mathbf{p}_\perp|$. Hence with a good accuracy we can expand the square root in the integrand in powers of $|\mathbf{p}|/p_0$. Taking into account the linear and quadratic terms in the square root we get

$$[(p_z+p_0)^2 + \mathbf{p}_\perp^2]^{1/2} \approx p_0 + p_z + \mathbf{p}_\perp^2/2p_0 \tag{2.20}$$

Then the integral over $d^3\mathbf{p}$ can be calculated as the product of integrals over $d^2\mathbf{p}_\perp$ and dp_z and the calculation is analogous to that in Eq. (2.8). The result of the calculation is

$$\psi(\mathbf{r},t) = [\pi^{3/4}ab^{1/2}(1+\frac{i\hbar ct}{p_0a^2})]^{-1}exp[\frac{i}{\hbar}(\mathbf{p}_0\mathbf{r} - p_0ct)]$$
$$exp[-\frac{(\mathbf{r}_\perp - \mathbf{r}_{0\perp})^2(1-\frac{i\hbar ct}{p_0a^2})}{2a^2(1+\frac{\hbar^2c^2t^2}{p_0^2a^4})} - \frac{(z-z_0-ct)^2}{2b^2}] \tag{2.21}$$

This result shows that the wave packet describing an ultrarelativistic particle (including a photon) is moving along the classical trajectory $z(t) = z_0 + ct$, in the longitudinal direction there is no spreading while in transversal directions spreading is characterized by the function

$$a(t) = a(1+\frac{\hbar^2c^2t^2}{p_0^2a^4})^{1/2} \tag{2.22}$$

The characteristic time of spreading can be defined as $t_* = p_0a^2/\hbar c$. The fact that $t_* \to \infty$ in the formal limit $\hbar \to 0$ shows that in relativistic case WPS also is a pure quantum phenomenon (see the end of Sec. 2.3). From the formal point of view the result for t_* is the same as in nonrelativistic theory but m should be replaced by E/c^2 where E is the energy of the ultrarelativistic particle. This fact could be expected since, as noted above, it is reasonable to think that spreading in directions perpendicular to the particle momentum is similar to that in standard nonrelativistic quantum mechanics. However, in the ultrarelativistic case spreading takes place only in this direction. If $t \gg t_*$ the transversal width of the packet is $a(t) = \hbar ct/p_0a$. Hence the speed of spreading in the perpendicular direction is $v_* = \hbar c/p_0a$.

2.6 Geometrical optics

The relation between quantum and classical electrodynamics is well-known and is described in textbooks (see e.g. Ref. [12]). As already noted, classical electromagnetic field consists of many photons and in classical electrodynamics the photons are not described individually. Instead, classical electromagnetic field is described by field strengths which represent mean characteristics of a large set of photons. For constructing the field strengths one can use the photon wave functions $\chi(\mathbf{p}, t)$ or $\psi(\mathbf{r}, t)$ where E is replaced by $\hbar\omega$ and $\mathbf{p}$ is replaced by $\hbar\mathbf{k}$. In this connection it is interesting to note that since ω is a classical quantity used for describing a classical electromagnetic field, the photon is a pure quantum particle since its energy disappears in the formal limit $\hbar \to 0$. Even this fact shows that the photon cannot be treated as a classical particle and the effect of WPS for the photon cannot be neglected.

With the above replacements the functions χ and ψ do not contain any dependence on $\hbar$ (note that the normalization factor $\hbar^{-3/2}$ in $\chi(\mathbf{k}, t)$ will disappear since the normalization integral for $\chi(\mathbf{k}, t)$ is now over $d^3\mathbf{k}$, not $d^3\mathbf{p}$). The quantities ω and $\mathbf{k}$ are now treated, respectively, as the frequency and the wave vector of the classical electromagnetic field and the functions $\chi(\mathbf{k}, t)$ and $\psi(\mathbf{r}, t)$ are interpreted not such that they describe probabilities for a single photon but such that they describe classical electromagnetic field and $\mathbf{E}(\mathbf{r}, t)$ and $\mathbf{B}(\mathbf{r}, t)$ can be constructed from these functions as described in textbooks on QED (see e.g. Ref. [12]).

An additional argument in favor of the choice of $\psi(\mathbf{r}, t)$ as the coordinate photon wave function is that in classical electrodynamics the quantities $\mathbf{E}(\mathbf{r}, t)$ and $\mathbf{B}(\mathbf{r}, t)$ for the free field should satisfy the wave equation $\partial^2\mathbf{E}/c^2\partial t^2 = \Delta\mathbf{E}$ and analogously for $\mathbf{B}(\mathbf{r}, t)$. Hence if $\mathbf{E}(\mathbf{r}, t)$ and $\mathbf{B}(\mathbf{r}, t)$ are constructed from $\psi(\mathbf{r}, t)$ as described in textbooks (see e.g. Ref. [12]), they will satisfy the wave equation since, as follows from Eqs. (2.7,2.17,2.18), $\psi(\mathbf{r}, t)$ also satisfies this equation.

The geometrical optics approximation implies that if $\mathbf{k}_0$ and $\mathbf{r}_0$ are the mean values of the wave vector and the spatial radius vector for a wave packet describing the electromagnetic wave then the uncertainties Δk and Δr, which are the mean values of $|\mathbf{k} - \mathbf{k}_0|$ and $|\mathbf{r} - \mathbf{r}_0|$, respectively, should satisfy the requirements $\Delta k \ll |\mathbf{k}_0|$ and $\Delta r \ll |\mathbf{r}_0|$. Analogously, in full analogy with the derivation of Eq. (2.2), one can show that for each $j = 1, 2, 3$ the uncertainties of the corresponding projections of the vectors $\mathbf{k}$ and $\mathbf{r}$ satisfy the requirement $\Delta k_j \Delta r_j \geq 1/2$ (see e.g. Ref. [3]). In particular, an electromagnetic wave satisfies the approximation of geometrical optics in the greatest possible extent if $\Delta k \Delta r$ is of the order of unity.

The above discussion confirms what has been mentioned in Sec. 2.1 that *the effect of WPS in transverse directions takes place not only in quantum theory but even in classical electrodynamics.* Indeed, since the function $\psi(\mathbf{r}, t)$ satisfies the classical wave equation, the above consideration can be also treated as an example showing that *even for a free wave packet in classical electrodynamics the WPS effect is inevitable.* In the language of classical waves the parameters of spreading can be characterized by the function $a(t)$ (see Eq. (2.22)) and the quantities t_* and v_* (see the end of the preceding section) such that in terms of the wave length $\lambda = 2\pi c/\omega_0$

$$a(t) = a(1+\frac{\lambda^2 c^2 t^2}{4\pi^2 a^4})^{1/2}, \quad t_* = \frac{2\pi a^2}{\lambda c}, \quad v_* = \frac{\lambda c}{2\pi a} \tag{2.23}$$

The last expression can be treated such that if $\lambda \ll a$ then the momentum has the angular uncertainty of the order of $\alpha = \lambda/(2\pi a)$. This result is natural from the following consideration. Let the mean value of the momentum be directed along the z-axis and the uncertainty of the transverse component of the momentum be $\Delta p_\perp$. Then $\Delta p_\perp$ is of the order of $\hbar/a$, $\lambda = 2\pi\hbar/p_0$ and hence α is of the order of $\Delta p_\perp/p_0 \approx \lambda/(2\pi a)$. This is analogous to the well-known result in classical optics that the best angular resolution of a telescope with the dimension d is of the order of λ/d. Another well-known result of classical optics is that if a wave encounters an obstacle having the dimension d then the direction of the wave diverges by the angle of the order of λ/d.

The inevitability of WPS for a free wave packet in classical electrodynamics is obvious from the following consideration. Suppose that a classical wave packet does not have a definite value of the momentum. Then if a is the initial width of the packet in directions perpendicular to the mean momentum, one might expect that the width will grow as $a(t) = a + \alpha ct$ and for large values of t, $a(t) \approx \alpha ct$. As follows from Eq. (2.23), if $t \gg t_*$ then indeed $a(t) \approx \alpha ct$. In standard quantum theory we have the same result because the coordinate and momentum wave functions are related to each other by the same Fourier transform as the coordinate and $\mathbf{k}$ distributions in classical electrodynamics.

The quantity $N_{||} = b/\lambda$ shows how many oscillations the oscillating exponent in Eq. (2.21) makes in the region where the wave function or the amplitude of the classical wave is significantly different from zero. As noted in Sec. 2.2, for the validity of semiclassical approximation this quantity should be very large. In nonrelativistic quantum mechanics a and b are of the same order and hence the same can be said about the quantity $N_\perp = a/\lambda$. As noted above, in the case of the photon we don't know the relation between a and b. In terms of the quantity $N_\perp$ we can rewrite the expressions for t_* and v_* in Eq. (2.23) as

$$t_* = 2\pi N_\perp^2 T, \quad v_* = \frac{c}{2\pi N_\perp} \tag{2.24}$$

where T is the period of the classical wave. Hence the accuracy of semiclassical approximation (or the geometrical optics approximation in classical electrodynamics) increases with the increase of $N_\perp$.

In Ref. [61] the problem of WPS for classical electromagnetic waves has been discussed in the Fresnel approximation (i.e. in the approximation of geometrical optics) for a two-dimensional wave packet. Equation (25) of Ref. [61] is a special case of Eq. (2.20) and the author of Ref. [61] shows that, in his model the wave packet spreads out in the direction perpendicular to the group velocity of the packet. As noted at the end of the preceding section, in the ultrarelativistic case the function $a(t)$ is given by the same expression as in the nonrelativistic case but m is replaced by E/c^2. Hence if the results of the preceding section are reformulated in terms of classical waves then m should be replaced by $\hbar\omega_0/c^2$ and this fact has been pointed out in Ref. [61].

2.7 Wave packet width paradox

We now consider the following important question. We assume that a classical wave packet is a collection of photons. Let a_{cl} be the quantity a for the classical packet and a_{ph} be a typical value of a for the photons. What is the relation between a_{cl} and a_{ph}?

My observation is that physicists answer this question in different ways. Quantum physicists usually say that in typical situations $a_{ph} \ll a_{cl}$ because a_{cl} is of macroscopic size while in semiclassical approximation the quantity a_{ph} for each photon can be treated as the size of the region where the photon has been created. On the other hand, classical physicists usually say that $a_{ph} \gg a_{cl}$ and the motivation is as follows.

Consider a decomposition of some component of classical electromagnetic field into the Fourier series:

$$A(x) = \sum_{\sigma} \int [a(\mathbf{p},\sigma)u(\mathbf{p},\sigma)exp(-ipx) + a(\mathbf{p},\sigma)^* u(\mathbf{p},\sigma)^* exp(ipx)] d^3\mathbf{p} \qquad (2.25)$$

where σ is the polarization, x and p are the four-vectors such that $x = (ct, \mathbf{x})$ and $p = (|\mathbf{p}|c, \mathbf{p})$, the functions $a(\mathbf{p},\sigma)$ are the same for all the components, the functions $u(\mathbf{p},\sigma)$ depend on the component and * is used to denote the complex conjugation. Then photons arise as a result of quantization when $a(\mathbf{p},\sigma)$ and $a(\mathbf{p},\sigma)^*$ are understood not as usual function but as operators of annihilation and creation of the photon with the quantum numbers $(\mathbf{p},\sigma)$ and * is now understood as Hermitian conjugation. Hence the photon is described by a plane wave which has the same magnitude in all points of the space. In other words, a_{ph} is infinitely large and a finite width of the classical wave packet arises as a result of interference of different plane waves.

The above definition of the photon has at least the following inconsistency. If the photon is treated as a particle then its wave function should be normalizable while the plane wave is not normalizable. In textbooks this problem is often circumvented by saying that we consider our system in a finite box. Then the spectrum of momenta becomes finite and instead of Eq. (2.25) one can write

$$A(x) = \sum_{\sigma} \sum_{j} [a(\mathbf{p}_j,\sigma)u(\mathbf{p}_j,\sigma)exp(-ip_jx) + a(\mathbf{p}_j,\sigma)^* u(\mathbf{p}_j,\sigma)^* exp(ip_jx)] \qquad (2.26)$$

where j enumerates the points of the momentum spectrum.

One can now describe quantum electromagnetic field by states in the Fock space where the vacuum vector Φ_0 satisfies the condition $a(\mathbf{p}_j,\sigma)\Phi_0=0$, $||\Phi_0||=1$ and the operators commute as

$$[a(\mathbf{p}_i,\sigma_k),a(\mathbf{p}_j,\sigma_l)]=[a(\mathbf{p}_i,\sigma_k)^*,a(\mathbf{p}_j,\sigma_l)^*]=0,\quad [a(\mathbf{p}_i,\sigma_k),a(\mathbf{p}_j,\sigma_l)^*]=\delta_{ij}\delta_{kl} \tag{2.27}$$

Then any state can be written as

$$\Psi=\sum_{n=0}^{\infty}\sum_{\sigma_1...\sigma_n}\sum_{\mathbf{p}_1,...\mathbf{p}_n}\chi(\mathbf{p}_1,\sigma_1,...\mathbf{p}_n,\sigma_n)a(\mathbf{p}_1,\sigma_1)^*\cdots a(\mathbf{p}_n,\sigma_n)^*\Phi_0 \tag{2.28}$$

Classical states are understood such that although the number of photons is large, it is much less than the number of possible momenta and in Eq. (2.28) all the photons have different momenta (this is analogous to the situation in classical statistics where mean numbers of particles in each state are much less than unity). Then it is not important whether the operators (a,a^*) commute or anticommute. However, according to the Pauli theorem on the spin-statistics connection [11], they should commute and this allows the existence of coherent states where many photons have the same quantum numbers. Such states can be created in lasers and they are not described by classical electrodynamics. In the next section we consider position operator for coherent states while in this section we consider only quantum description of states close to classical.

The next problem is that one should take into account that in standard theory the photon momentum spectrum is continuous. Then the above construction can be generalized as follows. The vacuum state Φ_0 satisfies the same conditions $||\Phi_0||=1$ and $a(\mathbf{p},\sigma)\Phi_0=0$ while the operators (a,a^*) satisfy the following commutation relations

$$[a(\mathbf{p},\sigma),a(\mathbf{p}',\sigma')]=[a(\mathbf{p},\sigma)^*,a(\mathbf{p}',\sigma')^*]=0\quad [a(\mathbf{p},\sigma),a(\mathbf{p}',\sigma')^*]=\delta^{(3)}(\mathbf{p}-\mathbf{p}')\delta_{\sigma\sigma'} \tag{2.29}$$

Then a general quantum state can be written as

$$\Psi=\sum_{n=0}^{\infty}\sum_{\sigma_1...\sigma_n}\int ...\int \chi(\mathbf{p}_1,\sigma_1,...\mathbf{p}_n,\sigma_n)a(\mathbf{p}_1,\sigma_1)^*\cdots a(\mathbf{p}_n,\sigma_n)^*d^3\mathbf{p}_1\cdots d^3\mathbf{p}_n\Phi_0 \tag{2.30}$$

In the approximation when a classical wave packet is understood as a collection of independent photons (see the discussion in Sec. 10), the state of this packet has the form

$$\Psi=\sum_{n=0}^{\infty}c_n\prod_{j=1}^{n}\{\sum_{\sigma_j}\int \chi_j(\mathbf{p}_j,\sigma_j)a(\mathbf{p}_j,\sigma_j)^*d^3\mathbf{p}_j\}\Phi_0 \tag{2.31}$$

where χ_j is the wave function of the jth photon and intersections of supports of wave functions of different photons can be neglected. This is an analog of the above situation with the discrete case where it is assumed that different photons in a classical wave packet have different momenta. In other words, while the wave function of each photon can be treated as an interference of plane waves, different photons can interfere only in coherent states but not in classical wave packets.

We now describe a well-known generalization of the results on IRs of the Poincare algebra to the description in the Fock space (see e.g. Ref. [37] for details). If A is an operator in the space of the photon IR then a generalization of this operator to the case of the Fock space can be constructed as follows. Any operator in the space of IR can be represented as an integral operator acting on the wave function as

$$A\chi(\mathbf{p},\sigma) = \sum_{\sigma'} \int A(\mathbf{p},\sigma,\mathbf{p}',\sigma')\chi(\mathbf{p}',\sigma')d^3\mathbf{p}' \tag{2.32}$$

For example, if $\mathbf{A}\chi(\mathbf{p},\sigma) = \partial\chi(\mathbf{p},\sigma)/\partial\mathbf{p}$ then $\mathbf{A}$ is the integral operator with the kernel

$$\mathbf{A}(\mathbf{p},\sigma,\mathbf{p}',\sigma') = \frac{\partial\delta^{(3)}(\mathbf{p}-\mathbf{p}')}{\partial\mathbf{p}}\delta_{\sigma\sigma'}$$

We now require that if the action of the operator A in the space of IR is defined by Eq. (2.32) then in the case of the Fock space this action is defined as

$$A = \sum_{\sigma\sigma'} \int A(\mathbf{p},\sigma,\mathbf{p}',\sigma')a(\mathbf{p},\sigma)^* a(\mathbf{p}',\sigma')d^3\mathbf{p}d^3\mathbf{p}' \tag{2.33}$$

Then it is easy to verify that if A, B and C are operators in the space of IR satisfying the commutation relation $[A, B] = C$ then the generalizations of these operators in the Fock space satisfy the same commutation relation. It is also easy to verify that the operators generalized to the action in the Fock space in such a way are additive, i.e. for a system of n photons they are sums of the corresponding single-particle operators. In particular, the energy of the n-photon system is a sum of the energies of the photons in the system and analogously for the other representation operators of the Poincare algebra - momenta, angular momenta and Lorentz boosts.

We are interested in calculating mean values of different combinations of the momentum operator. Since this operator does not act over spin variables, we will drop such variables in the (a, a^*) operators and in the functions χ_j. Then the explicit form of the momentum operator is $\mathbf{P} = \int \mathbf{p}a(\mathbf{p})^* a(\mathbf{p})d^3\mathbf{p}$. Since this operator does not change the number of photons, the mean values can be independently calculated in each subspace where the number of photons is N.

Suppose that the momentum of each photon is approximately directed along the z-axis and the quantity p_0 for each photon approximately equals $2\pi\hbar/\lambda$. If $\Delta p_\perp$ is a typical uncertainty of the transversal component of the momentum for the photons then a typical value of the angular uncertainty for the photons is $\alpha_{ph} =$

$\Delta p_\perp/p_0 \approx \lambda/(2\pi a_{ph})$. The total momentum of the classical wave packet consisting of N photons is a sum of the photon momenta: $\mathbf{P} = \sum_{i=1}^N \mathbf{p}^{(i)}$. Suppose that the mean value of $\mathbf{P}$ is directed along the z-axis and its magnitude P_0 is such that $P_0 \approx Np_0$. The uncertainty of the x component of $\mathbf{P}$ is $\Delta P_x = \overline{P_x^2}^{1/2}$ where

$$\overline{P_x^2} = \sum_{i=1}^N \overline{(p_x^{(i)})^2} + \sum_{i\neq j;i,j=1}^N \overline{p_x^{(i)} p_x^{(j)}}$$

Then in the approximation of independent photons (see the remarks after Eq. (2.31))

$$\overline{P_x^2} = \sum_{i=1}^N \overline{(p_x^{(i)})^2} + \sum_{i\neq j;i,j=1}^N \overline{p_x^{(i)}} \cdot \overline{p_x^{(j)}} = \sum_{i=1}^N [\overline{(p_x^{(i)})^2} - \overline{p_x^{(i)}}^2] = \sum_{i=1}^N (\Delta p_\perp^{(i)})^2$$

where we have taken into account that $\overline{P_x} = \sum_{i=1}^N \overline{p_x^{(i)}} = 0$.

As a consequence, if typical values of $\Delta p_\perp^{(i)}$ have the the same order of magnitude equal to $\Delta p_\perp$ then $\Delta P_\perp \approx N^{1/2}\Delta p_\perp$ and the angular divergence of the classical vave packet is

$$\alpha_{cl} = \Delta P_\perp/P_0 \approx \Delta p_\perp/(p_0 N^{1/2}) = \alpha_{ph}/N^{1/2}$$

Since the classical wave packet is described by the same wave equation as the photon wave function, its angular divergence can be expressed in terms of the parameters λ and a_{cl} such that $\alpha_{cl} = \lambda/(2\pi a_{cl})$. Hence $a_{cl} \approx N^{1/2} a_{ph}$ and we conclude that $a_{ph} \ll a_{cl}$.

Note that in this derivation no position operator has been used. Although the quantities λ and a_{ph} have the dimension of length, they are defined only from considering the photon in momentum space because, as noted in Sec. 1.2, for individual photons λ is understood only as $2\pi\hbar/p_0$, a_{ph} defines the width of the photon momentum wave function (see Eq. (2.17)) and is of the order of $\hbar/\Delta p_\perp$. As noted in Secs. 2.3 and 2.5, the momentum distribution does not depend on time and hence the result $a_{ph} \ll a_{cl}$ does not depend on time too. If photons in a classical wave packet could be treated as (almost) pointlike particles then photons do not experience WPS while the WPS effect for a classical wave packet is a consequence of the fact that different photons in the packet have different momenta.

However, in standard quantum theory this scenario does not take place for the following reason. Let $a_{cl}(t)$ be the quantity $a(t)$ for the classical wave packet and $a_{ph}(t)$ be a typical value of the quantity $a(t)$ for individual photons. With standard position operator the quantity $a_{ph}(t)$ is interpreted as the spatial width of the photon coordinate wave function in directions perpendicular to the photon momentum and this quantity is time dependent. As shown in Secs. 2.5 and 2.6, $a(0) = a$ but if $t \gg t_*$ then $a(t)$ is *inversely proportional* to a and the coefficient of proportionality is the same for the classical wave packet and individual photons (see Eq. (2.23)).

Hence *in standard quantum theory we have a paradox that after some period of time* $a_{ph}(t) \gg a_{cl}(t)$ i.e. individual photons in a classical wave packet spread out in a much greater extent than the wave packet as a whole. We call this situation the wave packet width (WPW) paradox (as noted above, different photons in a classical wave packet do not interfere with each other). The reason of the paradox is obvious: if the law that the angular divergence of a wave packet is of the order of λ/a is applied to both, a classical wave packet and photons comprising it then the paradox follows from the fact that the quantities a for the photons are much less than the quantity a for the classical wave packet. Note that in classical case the quantity a_{cl} does not have the meaning of $\hbar/\Delta P_{\perp}$ and λ is not equal to $2\pi\hbar/P_0$.

2.8 Wave packet spreading in coherent states

In textbooks on quantum optics the laser emission is described by the following model (see e.g. Ref. [62]). Consider a set of photons having the same momentum $\mathbf{p}$ and polarization σ and, by analogy with the discussion in the preceding section, suppose that the momentum spectrum is discrete. Consider a quantum superposition $\Psi = \sum_{n=0}^{\infty} c_n [a(\mathbf{p},\sigma)^*]^n \Phi_0$ where the coefficients c_n satisfy the condition that Ψ is an eigenstate of the annihilation operator $a(\mathbf{p},\sigma)$. Then the product of the coordinate and momentum uncertainties has the minimum possible value $\hbar/2$ and, as noted in Sec. 2.2, such a state is called coherent. However, the term coherent is sometimes used meaning that the state is a quantum superposition of many-photon states $[a(\mathbf{p},\sigma)^*]^n\Phi_0$.

In the above model it is not taken into account that (in standard theory) photons emitted by a laser can have only a continuous spectrum of momenta. Meanwhile for the WPS effect the width of the momentum distribution is important. In this section we consider a generalization of the above model where the fact that photons have a continuous spectrum of momenta is taken into account. This will make it possible to consider the WPS effect in coherent states.

In the above formalism coherent states can be defined as follows. We assume that all the photons in the state Eq. (2.30) have the same polarization. Hence for describing such states we can drop the quantum number σ in wave functions and a-operators. We also assume that all photons in coherent states have the same momentum distribution. These conditions can be satisfied by requiring that coherent states have the form

$$\Psi = \sum_{n=0}^{\infty} c_n [\int \chi(\mathbf{p}) a(\mathbf{p})^* d^3\mathbf{p}]^n \Phi_0 \tag{2.34}$$

where c_n are some coefficients. Finally, by analogy with the description of coherent states in standard textbooks on quantum optics one can require that they are eigenstates of the operator $\int a(\mathbf{p}) d^3\mathbf{p}$.

The dependence of the state Ψ in Eq. (2.34) on t is $\Psi(t) = exp(-iEt/\hbar)\Psi$

where, as follows from Eqs. (2.12) and (2.33), the action of the energy operator in the Fock space is $E = \int pca(\mathbf{p})^* a(\mathbf{p}) d^3\mathbf{p}$. Since $exp(iEt/\hbar)\Phi_0 = \Phi_0$, it readily follows from Eq. (2.29) that

$$\Psi(t) = \sum_{n=0}^{\infty} c_n [\int \chi(\mathbf{p}, t) a(\mathbf{p})^* d^3\mathbf{p}]^n \Phi_0 \qquad (2.35)$$

where the relation between $\chi(\mathbf{p}, t)$ and $\chi(\mathbf{p}) = \chi(\mathbf{p}, 0)$ is given by Eq. (2.18).

A problem arises how to define the position operator in the Fock space. If this operator is defined by analogy with the above construction then we get an unphysical result that each coordinate of the n-photon system as a whole is a sum of the corresponding coordinates of the photons in the system. This is an additional argument that the position operator is less fundamental than the representation operators of the Poincare algebra and its action should be defined from additional considerations. In textbooks on quantum optics the position operator for coherent states is usually defined by analogy with the position operator in nonrelativistic quantum mechanics for the harmonic oscillator problem. The motivation is as follows. If the energy levels $\hbar\omega(n + 1/2)$ of the harmonic oscillator are treated as states of n quanta with the energies $\hbar\omega$ then the harmonic oscillator problem can be described by the operators a and a^* which are expressed in terms of the one-dimensional position and momentum operators q and p as $a = (\omega q + ip)/(2\hbar\omega)^{1/2}$ and $a^* = (\omega q - ip)/(2\hbar\omega)^{1/2}$. However, as noted above, the model description of coherent states in those textbooks is one-dimensional because the continuous nature of the momentum spectrum is not taken into account. In addition, the above results on WPS give indications that the position operator in standard theory is not consistently defined. For all these reasons a problem arises whether the requirement that the state Ψ in Eq. (2.34) is an eigenvector of the operator $\int a(\mathbf{p}) d^3\mathbf{p}$ has a physical meaning. In what follows this requirement will not be used.

In nonrelativistic classical mechanics the radius vector of a system of n particles as a whole (the radius vector of the center of mass) is defined as $\mathbf{R} = (m_1\mathbf{r}_1 + ... + m_n\mathbf{r}_n)/(m_1 + ... + m_n)$ and in works on relativistic classical mechanics it is usually defined as $\mathbf{R} = (\epsilon_1(\mathbf{p}_1)\mathbf{r}_1 + ... + \epsilon_n(\mathbf{p}_n)\mathbf{r}_n)/(\epsilon_1(\mathbf{p}_1) + ... + \epsilon_n(\mathbf{p}_n))$ where $\epsilon_i(\mathbf{p}_i) = (m_i^2 + \mathbf{p}_i^2)^{1/2}$. Hence if all the particles have the same masses and momenta, $\mathbf{R} = (\mathbf{r}_1 + ... + \mathbf{r}_n)/n$.

These remarks make it reasonable to define the position operator for coherent states as follows. Let x_j be the jth component of the position operator in the space of IR and $A_j(\mathbf{p}, \mathbf{p}')$ be the kernel of this operator. Then in view of Eq. (2.33) the action of the operator X_j on the state $\Psi(t)$ in Eq. (2.34) can be defined as

$$X_j\Psi(t) = \sum_{n=1}^{\infty} \frac{c_n}{n} \int\int A_j(\mathbf{p}", \mathbf{p}') a(\mathbf{p}")^* a(\mathbf{p}') d^3\mathbf{p}" d^3\mathbf{p}' [\int \chi(\mathbf{p}, t) a(\mathbf{p})^* d^3\mathbf{p}]^n \Phi_0 \quad (2.36)$$

If $\overline{x_j}(t)$ and $\overline{x_j^2}(t)$ are the mean values of the operators x_j and x_j^2, respectively then as follows from the definition of the kernel of the operator x_j

$$\overline{x_j}(t) = \int\int \chi(\mathbf{p},t)^* A_j(\mathbf{p},\mathbf{p}')\chi(\mathbf{p}',t)d^3\mathbf{p}d^3\mathbf{p}'$$
$$\overline{x_j^2}(t) = \int\int\int \chi(\mathbf{p}",t)^* A_j(\mathbf{p},\mathbf{p}")^* A_j(\mathbf{p},\mathbf{p}')\chi(\mathbf{p}',t)d^3\mathbf{p}d^3\mathbf{p}"d^3\mathbf{p}' \quad (2.37)$$

and in the case of IR the uncertainty of the quantity x_j is $\Delta x_j(t) = [\overline{x_j^2}(t) - \overline{x_j}(t)^2]^{1/2}$. At the same time, if $\overline{X_j}(t)$ and $\overline{X_j^2}(t)$ are the mean values of the operators X_j and X_j^2, respectively then

$$\overline{X_j}(t) = (\Psi(t), X_j\Psi(t)), \quad \overline{X_j^2}(t) = (\Psi(t), X_j^2\Psi(t)) \quad (2.38)$$

and the uncertainty of the quantity X_j is $\Delta X_j(t) = [\overline{X_j^2}(t) - \overline{X_j}(t)^2]^{1/2}$. Our goal is to express $\Delta X_j(t)$ in terms of $\overline{x_j}(t)$, $\overline{x_j^2}(t)$ and $\Delta x_j(t)$.

If the function $\chi(\mathbf{p},t)$ is normalized to one (see Eq. (2.5)) then, as follows from Eq. (2.29), $||\Psi(t)|| = 1$ if

$$\sum_{n=0}^{\infty} n!|c_n|^2 = 1 \quad (2.39)$$

A direct calculation using Eqs. (2.29), (2.36), (2.37) and (2.38) gives

$$\overline{X_j}(t) = \overline{x_j}(t)\sum_{n=1}^{\infty} n!|c_n|^2$$
$$\overline{X_j^2}(t) = \sum_{n=1}^{\infty}(n-1)!|c_n|^2[\overline{x_j^2}(t) + (n-1)\overline{x_j}(t)^2] \quad (2.40)$$

It now follows from Eq. (2.39) and the definitions of the quantities $\Delta x_j(t)$ and $\Delta X_j(t)$ that

$$\Delta X_j(t)^2 = (1-|c_0|^2)|c_0|^2\overline{x_j}(t)^2 + \sum_{n=1}^{\infty}(n-1)!|c_n|^2\Delta x_j(t)^2 \quad (2.41)$$

Equation (2.41) is the key result of this section. It has been derived without using a specific choice of the single photon position operator. The consequence of this result is as follows. If the main contribution to the state $\Psi(t)$ in Eq. (2.35) is given by very large values of n then $|c_0|$ is very small and the first term in this expression can be neglected. Suppose that the main contribution is given by terms where n is of the order of $\bar{n}$. Then, as follows from Eqs. (2.39) and (2.41), $\Delta X_j(t)$ is of the order of $\Delta x_j(t)/\bar{n}^{1/2}$. This means that for coherent states where the main contribution is given by very large numbers of photons the effect of WPS is pronounced in a much less extent than for single photons.

2.9 Experimental consequences of WPS in standard theory

The problem of explaining the redshift phenomenon has a long history. Different competing approaches can be divided into two big sets which we call Theory A and Theory B. In Theory A the redshift has been originally explained as a manifestation of the Doppler effect but in recent years the cosmological and gravitational redshifts have been added to the consideration. In this theory the interaction of photons with the interstellar medium is treated as practically not important, i.e. it is assumed that with a good accuracy we can treat photons as propagating in the empty space. On the contrary, in Theory B, which is often called the tired-light theory, the interaction of photons with the interstellar medium is treated as a main reason for the redshift. At present the majority of physicists believe that Theory A explains the astronomical data better than Theory B. Even some physicists working on Theory B acknowledged that any sort of scattering of light would predict more blurring than is seen (see e.g. the article "Tired Light" in Wikipedia).

A problem arises whether or not WPS of the photon wave function is important for explaining the redshift. One might think that this effect is not important since a considerable WPS would also blur the images more than what is seen. However, as shown in the previous discussion, WPS is an inevitable consequence of standard quantum theory and moreover this effect also exists in classical electrodynamics. Hence it is not sufficient to just say that a considerable WPS is excluded by observations. One should try to estimate the importance of WPS and to understand whether our intuition is correct or not.

As follows from these remarks, in Theory A it is assumed that with a good accuracy we can treat photons as propagating in the empty space. It is also reasonable to expect (see the discussion in the next section) that photons from distant stars practically do not interact with each other. Hence the effect of WPS can be considered for each photon independently and the results of the preceding sections make it possible to understand what experimental consequences of WPS are.

A question arises what can be said about characteristics of photons coming to Earth from distance objects. Since wave lengths of such photons are typically much less than characteristic dimensions of obstacles one might think that the radiation of stars can be described in the geometrical optics approximation. As discussed in Sec. 2.6, this approximation is similar to semiclassical approximation in quantum theory. This poses a question whether this radiation can be approximately treated as a collection of photons moving along classical trajectories.

Consider, for example, the Lyman transition $2P \to 1S$ in the hydrogen atom on the Sun. In this case the mean energy of the photon is $E_0 = 10.2eV$, its wave length is $\lambda = 121.6nm$ and the lifetime is $\tau = 1.6 \cdot 10^{-9}s$. The phrase that the lifetime is τ is interpreted such that the uncertainty of the energy is $\hbar/\tau$, the uncertainty of the longitudinal momentum is $\hbar/c\tau$ and b is of the order of $c\tau \approx 0.48m$. In this case the photon has a very narrow energy distribution since the mean value of the momentum $p_0 = E_0/c$ satisfies the condition $p_0 b \gg \hbar$. At the same time, since the orbital angular momentum of the photon is a small quantity, the direction of the photon momentum cannot be semiclassical. Qualitative features of such situations can be described by the following model.

Suppose that the photon momentum wave function is spherically symmetric and has the form

$$\chi(p) = Cexp[-\frac{1}{2}(p-p_0)^2 b^2 - \frac{i}{\hbar}pr_0] \tag{2.42}$$

where C is a constant, and p is the magnitude of the momentum. Then the main contribution to the normalization integral is given by the region of p where $|p-p_0|$ is of the order of $\hbar/b$ and in this approximation the integration over p can be taken from $-\infty$ to ∞. As a result, the function normalized to one has the form

$$\chi(p) = \frac{b^{1/2}}{2\pi^{3/4}p_0}exp[-\frac{1}{2\hbar^2}(p-p_0)^2 b^2 - \frac{i}{\hbar}pr_0] \tag{2.43}$$

The dependence of this function on t is $\chi(p,t) = exp(-iE(p)t/\hbar)\chi(p)$ where $E(p) = pc$. Hence

$$\chi(p,t) = \frac{b^{1/2}}{2\pi^{3/4}p_0}exp[-\frac{1}{2\hbar^2}(p-p_0)^2 b^2 - \frac{i}{\hbar}pr_0(t)] \tag{2.44}$$

where $r_0(t) = r_0 + ct$.

The coordinate wave function is

$$\psi(\mathbf{r},t) = \frac{1}{(2\pi\hbar)^{3/2}}\int \chi(p,t)e^{i\mathbf{pr}/\hbar}d^3\mathbf{p} \tag{2.45}$$

Since $\chi(p,t)$ is spherically symmetric it is convenient to decompose $e^{i\mathbf{pr}/\hbar}$ as a sum of spherical harmonics and take into account that only the term corresponding to $l = 0$ contributes to the integral. This term is $j_0(pr/\hbar) = sin(pr/\hbar)/(pr/\hbar)$. Then the integral can be again taken from $-\infty$ to ∞ and the result is

$$\psi(r,t) = \frac{1}{2i\pi^{3/4}r_0(t)b^{1/2}}exp[-\frac{(r-r_0(t))^2}{2b^2} + \frac{i}{\hbar}p_0(r-r_0(t))] \tag{2.46}$$

We assume that $r_0(t) \gg b$ and hence the term with $exp[-(r+r_0(t))^2/2b^2]$ can be neglected and r in the denominator can be replaced by $r_0(t)$. As follows from the above results, the mean value of r is $r_0(t)$. If λ is defined as $\lambda = 2\pi\hbar/p_0$ then the requirement that $p_0 b \gg \hbar$ implies that $b \gg \lambda$. The conditions $p_0 b/\hbar \gg 1$ and $r_0(t) \gg b$ imply that the radial part of the photon state is semiclassical while the angular part is obviously strongly nonclassical.

Suppose that we want to detect the photon inside the volume V where the coordinates are $x \in [-d_x, d_x]$, $y \in [-d_y, d_y]$, $z \in [r_0(t) - d_z, r_0(t) + d_z]$. Let $g(\mathbf{r})$

be the characteristic function of V, i.e. $g(\mathbf{r}) = 1$ when $\mathbf{r} \in V$ and $g(\mathbf{r}) = 0$ otherwise. Let $\mathcal{P}$ be the projector acting on wave functions as $\mathcal{P}\psi(\mathbf{r}) = g(\mathbf{r})\psi(\mathbf{r})$. Then

$$\mathcal{P}\psi(r,t) = \frac{1}{2i\pi^{3/4}r_0(t)b^{1/2}}g(\mathbf{r})exp[-\frac{(r-r_0(t))^2}{2b^2} + \frac{i}{\hbar}p_0(r-r_0(t))] \tag{2.47}$$

Assume that $r_0(t) \gg d_x, d_y$. Then $r - r_0(t) \approx z - r_0(t) + (x^2+y^2)/2r_0(t)$. We also assume that $r_0(t)$ is so large then $r_0(t)\lambda \gg (d_x^2 + d_y^2)$. Then

$$\mathcal{P}\psi(r,t) \approx \frac{1}{2i\pi^{3/4}r_0(t)b^{1/2}}g(\mathbf{r})exp[-\frac{(z-r_0(t))^2}{2b^2} + \frac{i}{\hbar}p_0(z-r_0(t))] \tag{2.48}$$

We also assume that $d_z \gg b$. Then a simple calculation shows that

$$||\mathcal{P}\psi(r,t)||^2 = \frac{S}{4\pi r_0(t)^2} \tag{2.49}$$

where $S = 4d_x d_y$ is the area of the cross section of V by the plane $z = r_0(t)$. The meaning of Eq. (2.49) is obvious: $||\mathcal{P}\psi(r,t)||^2$ is the ratio of the cross section to the area of the sphere with the radius $r_0(t)$.

Let us now calculate the momentum distribution in the function $\mathcal{P}\psi(r,t)$. This distribution is defined as

$$\tilde{\chi}(\mathbf{p}) = \frac{1}{(2\pi\hbar)^{3/2}}\int [\mathcal{P}\psi(r,t)]e^{-i\mathbf{pr}/\hbar}d^3\mathbf{r} \tag{2.50}$$

As follows from Eq. (2.47)

$$\tilde{\chi}(\mathbf{p}) = A(t)exp[-\frac{1}{2\hbar^2}(p_z-p_0)^2b^2]j_0(p_x d_x/\hbar)j_0(p_y d_y/\hbar) \tag{2.51}$$

where $A(t)$ is a function of t. This result is similar to the well-known result in optics that the best angular resolution is of the order of λ/d where d is the dimension of the optical device (see e.g. textbooks [62]). As noted in Sec. 2.1 the reason of the similarity is that in quantum theory the coordinate and momentum representations are related to each other by the Fourier transform by analogy with classical electrodynamics. Note also that since the fall off of the function $j_0(x) = sinx/x$ is not rapid enough when x increases, in the case when many photons are detected, a considerable part of them might be detected in the angular range much greater than λ/d.

Let L be the distance to a pointlike source of spherically symmetric photons. From geometrical consideration one might expect that photons from this source will be detected in the angular range of the order of d/L. This quantity does not depend on λ while the quantity λ/d does not depend on L. Therefore the result given by Eq. (2.51) is counterintuitive. It is shown in Sec. 2.13 that, in contrast to the standard result λ/d obtained with the Fourier transform, our position operator indeed predicts the angular resolution of the order of d/L.

If R is the radius of a star then one might expect that the star will be visible in the angular range $(R+d)/L \approx R/L$. Hence the standard result predicts that if $\lambda/d \geq R/L$ then the image of the star will be blurred. The experimental verification of this prediction is problematic since the quantities R/L are very small and at present star radii cannot be measured directly. Conclusions about them are made from the data on luminosity and temperature assuming that the major part of the radiation from stars comes not from transitions between atomic levels but from processes which can be approximately described as a blackbody radiation.

A theoretical model describing blackbody radiation (see e.g. Ref. [63]) is such that photons are treated as an ideal Bose gas weakly interacting with matter and such that typical photon energies are not close to energies of absorption lines for that matter (hence the energy spectrum of photons is almost continuous). It is also assumed that the photons are distributed over states with definite values of momenta. With these assumptions one can derive the famous Planck formula for the spectral distribution of the blackbody radiation (this formula is treated as marking the beginning of quantum theory). When the photons leave the black body, their distribution in the phase space can be described by the Liouville theorem; in particular it implies that the photons are moving along classical trajectories.

Although the blackbody model is not ideal, numerous experimental data indicate that it works with a good accuracy. One of the arguments that the major part of the radiation consists of semiclassical photons is that the data on deflection of light by the Sun are described in the eikonal approximation which shows that the light from stars consists mainly of photons approximately moving along classical trajectories.

If we accept those arguments that the main part of photons emitted by stars can be qualitatively described in the formalism considered in Sec. 2.5. In that case we cannot estimate the quantity b as above and it is not clear what criteria can be used for estimating the quantity a.

The estimation $a \approx b \approx 0.48m$ seems to be very favorable since one might expect that the value of a is of atomic size, i.e. much less than $0.48m$. With this estimation for yellow light (with $\lambda = 580nm$) $N_\perp = a/\lambda \approx 8 \cdot 10^5$. So the value of $N_\perp$ is rather large and in view of Eq. (2.24) one might think that the effect of spreading is not important. However, this is not the case because, as follows from Eq. (2.24), $t_* \approx 0.008s$. Since the distance between the Sun and the Earth is approximately $t = 8$ light minutes and this time is much greater than t_*, the value of $a(t)$ (which can be called the half-width of the wave packet) when the packet arrives to the Earth is $v_* t \approx 28km$. In this case standard geometrical interpretation obviously does not apply. In addition, if we assume that the initial value of a is of the order of several wave lengths then the value of $N_\perp$ is much less and the width of the wave packet coming to the Earth is much greater. An analogous estimation shows that even in the favorable scenario the half-width of the wave packet coming to the Earth from Sirius will be approximately equal to $15 \cdot 10^6 km$ but in less favorable situations the half-width will be much greater. Hence we come to the conclusion that even in favorable scenarios the assumption that photons are moving along classical trajectories does not apply and a problem arises whether or not this situation is in agreement with experiment.

For illustration we consider the following example. Let the Earth be at point A and the center of Sirius be at point B. Suppose for simplicity that the Earth is a pointlike particle. Suppose that Sirius emitted a photon such that its wave function in momentum space has a narrow distribution around the mean value directed not along BA but along BC such that the angle between BA and BC is α. As noted in Sec. 2.5, there is no WPS in momentum space but, as follows from Eq. (2.23), the function $a(t)$ defining the mean value of the radius of the coordinate photon wave function in perpendicular directions is a rapidly growing function of t. Let us assume for simplicity that $\alpha \ll 1$. Then if L is the length of AB, the distance from A to BC is approximately $d = L\alpha$. So if this photon is treated as a point moving along the classical trajectory then the observer on the Earth will not see the photon. Let us now take into account the effect of WPS in directions perpendicular to the photon momentum. The front of the photon wave function passes the Earth when $t \approx t_1 = L/c$. As follows from Eq. (2.23) and the definition of the quantity $N_\perp$, if $t_1 \gg t_*$ then $a(t_1) = L/(2\pi N_\perp)$. If $a(t_1)$ is of the order of d or greater and we look in the direction AD such that AD is antiparallel to BC then there is a nonzero probability that we will detect this photon. So we can see photons coming from Sirius in the angular range which is of the order of $a(t_1)/L$. If R is the radius of Sirius and $a(t_1)$ is of the order of R or greater, the image of Sirius will be blurred. As noted above, a very optimistic estimation of $a(t_1)$ is $15 \cdot 10^6 km$. However, one can expect that a more realistic value of $N_\perp$ is not so large and then the estimation of $a(t_1)$ gives a much greater value. Since $R = 1.1 \cdot 10^6 km$ this means that the image of Sirius will be extremely blurred. Moreover, in the above angular range we can detect photons emitted not only by Sirius but also by other objects. Since the distance to Sirius is "only" 8.6 light years, for the majority of stars the effect of WPS will be pronounced even to a much greater extent. So if WPS is considerable then we will see not separate stars but an almost continuous background from many objects.

In the case of planets it is believed that we see a light reflected according to the laws of geometrical optics. Therefore photons of this light are in wave packet states and WPS for them can be estimated by using Eq. (2.23). The effect of blurring depends on the relation between the radii of planets and the corresponding quantities $a(t_1) = L/(2\pi N_\perp)$. Then it is obvious that if $N_\perp$ is not very large then even the images of planets will be blurred.

In the infrared and radio astronomy wave lengths are much greater than in the optical region but typical values of a_{ph} are expected to be much greater. As a consequence, predictions of standard quantum theory on blurring of astronomical images are expected to be qualitatively the same as in the optical region.

In the case of gamma-ray bursts (GRBs) wave lengths are much less than in the optical region but this is outweighed by the facts that, according to the present understanding of the GRB phenomenon (see e.g. Ref. [64]), gamma quanta created in GRBs typically travel to Earth for billions of years and typical values of a_{ph} are expected to be much less than in the optical region. The location of sources of GBRs are determined with a good accuracy and the data can be explained only assuming that the gamma quanta are focused into narrow jets (i.e. GRBs are not spherically symmetric) which are observable when Earth lies along the path of those jets. However, in view of the above discussion, the results on WPS predicted by standard quantum theory are fully incompatible with the data on GRBs.

A striking example illustrating the problem with the WPS effect follows. The phenomenon of the relic (CMB) radiation is treated as a case where the approximation of the blackbody radiation works with a very high accuracy. As noted above, photons emitted in this radiation are treated as moving along classical trajectories i.e. that they are in wave packet states. Since their wave lengths are much greater than wave lengths in the optical region and the time of their travel to Earth is several billions of years, the quantity $a(t_1)$ should be so large that no anisotropy of CMB should be observable. Meanwhile the anisotropy is observable and in the literature different mechanisms of the anisotropy are discussed (see e.g. Ref. [65]). However, the effect of WPS is not discussed.

On the other hand, the effect of WPS is important only if a particle travels a rather long distance. Hence one might expect that in experiments on the Earth this effect is negligible. Indeed, one might expect that in typical experiments on the Earth the quantity t_1 is so small that $a(t_1)$ is much less than the size of any macroscopic source of light. However, a conclusion that the effect of WPS is negligible for any experiment on the Earth might be premature.

As an example, consider the case of protons in the LHC accelerator. According to Ref. [66], protons in the LHC ring injected at the energy $E = 450GeV$ should be accelerated to the energy $E = 7TeV$ within one minute during which the protons will turn around the $27km$ ring approximately 674729 times. Hence the length of the proton path is of the order of $18 \cdot 10^6 km$. The protons cannot be treated as free particles since they are accelerated by strong magnets. A problem of how the width of the proton wave function behaves in the presence of strong electromagnetic field is very complicated and the solution of the problem is not known yet. It is always assumed that the WPS effect for the protons can be neglected. We will consider a model problem of the WPS for a free proton which moves for $t_1 = 1min$ with the energy in the range $[0.45, 7]TeV$.

In nuclear physics the size of the proton is usually assumed to be a quantity of the order of $10^{-13}cm$. Therefore for estimations we take $a = 10^{-13}cm$. Then the quantity t_* defined after Eq. (2.22) is not greater than $10^{-19}s$, i.e. $t_* \ll t_1$. Hence, as follows from Eq. (2.22), the quantity $a(t_1)$ is of the order of $500km$ if $E = 7TeV$ and in the case when $E = 450GeV$ this quantity is by a factor of $7/0.45 \approx 15.6$ greater. This fully unrealistic result cannot be treated as a paradox since, as noted above, the protons in the LHC ring are not free. Nevertheless it shows that a problem of what standard theory predicts on the width of proton wave functions in the LHC ring is far from being obvious.

Consider now WPS effects for radio wave photons. In radiolocation it is important that a beam from a directional antenna has a narrow angular distribution and a narrow distribution of wave lengths. Hence photons from the beam can be treated as (approximately) moving along classical trajectories. This makes it possible to communicate even with very distant space probes. For this purpose a set of radio telescopes can be used but for simplicity we consider a model where signals from a space probe are received by one radio telescope having the diameter D of the dish.

The Cassini spacecraft can transmit to Earth at three radio wavelengths: 14cm, 4cm and 1cm [67]. A radio telescope on Earth can determine the position of Cassini with a good accuracy if it detects photons having momenta in the angular range of the order of D/L where L is the distance to Cassini. The main idea of using a system of radiotelescopes is to increase the effective value of D. As a consequence of the fact that the radio signal sent from Cassini has an angular divergence which is much greater than D/L, only a small part of photons in the signal can be detected.

Consider first the problem on classical level. For the quantity $a = a_{cl}$ we take the value of $1m$ which is of the order of the radius of the Cassini antenna. If $\alpha = \lambda/(2\pi a)$ and $L(t)$ is the length of the classical path then, as follows from Eq. (2.23), $a_{cl}(t) \approx L(t)\alpha$. As a result, even for $\lambda = 1cm$ we have $a_{cl}(t) \approx 1.6 \cdot 10^6 km$. Hence if the photons in the beam are treated as (approximately) pointlike particles, one might expect that only a $[D/a_{cl}(t)]^2$ part of the photons can be detected.

Consider now the problem on quantum level. The condition $t \gg t_*$ is satisfied for both, the classical and quantum problems. Then, as follows from Eq. (2.23), $a_{ph}(t) = a_{cl}(t)a_{cl}/a_{ph}$, i.e. the quantity $a_{ph}(t)$ is typically greater than $a_{cl}(t)$ and in Sec. 2.7 this effect is called the WPW paradox. The fact that only photons in the angular range D/L can be detected can be described by projecting the states $\chi = \chi(\mathbf{p}, t)$ (see Eqs. (2.17), and (2.18)) onto the states $\chi_1 = \mathcal{P}\chi$ where $\chi_1(\mathbf{p}, t) = \rho(\mathbf{p})\chi(\mathbf{p}, t)$ and the form factor $\rho(\mathbf{p})$ is significant only if $\mathbf{p}$ is in the needed angular range. We choose $\rho(\mathbf{p}) = exp(-\mathbf{p}_\perp^2 a_1^2/2\hbar^2)$ where a_1 is of the order of $\hbar L/(p_0 D)$. Since $a_1 \gg a_{ph}$, it follows from Eqs. (2.17), and (2.18) that $||\mathcal{P}\chi||^2 = (a_{ph}/a_1)^2$. Then, as follows from Eq. (2.23), $(a_{ph}/a_1)^2$ is of the order of $[D/a_{ph}(t)]^2$ as expected and this quantity is typically much less than $[D/a_{cl}(t)]^2$. Hence the WPW paradox would make communications with space probes much more difficult.

Consider now the effect called Shapiro time delay. The meaning of the effect is as follows. An antenna on Earth sends a signal to Mercury, Venus or an interplanetary space probe and receives the reflected signal. If the path of the signal nearly grazes the Sun then the gravitational influence of the Sun deflects the path from a straight line. As a result, the path becomes longer by $S \approx 75km$ and the signals arrive with a delay $S/c \approx 250\mu s$. This effect is treated as the fourth test of GR and its theoretical consideration is based only on classical geometry. In particular, it is assumed that the radio signal is moving along the classical trajectory.

However, in standard quantum theory the length of the path has an uncertainty which can be defined as follows. As a consequence of WPS, the uncertainty of the path is

$$\Delta L(t) = [L(t)^2 + a(t)^2]^{1/2} - L(t) \approx a(t)^2/2L(t) = L(t)\alpha^2/2$$

In contrast to the previous example, this quantity is quadratic in α and one might think that it can be neglected. However, this is not the case. For example, in the first experiment on measuring the Shapiro delay [68] signals with the frequency 8GHz were sent by the MIT Haystack radar antenna [69] having the diameter $37m$. If we take for a_{ph} a very favorable value which equals the radius of the antenna then $\alpha^2 \approx 10^{-7}$. As a result, when the signal is sent to Venus, $\Delta L(t) \approx 25km$ but since a_{ph} is typically much less than a_{cl} then in view of the WPW paradox the value of $\Delta L(t)$ will be much greater. However, even the result $25km$ is incompatible with the fact that the accuracy of the experiment was of the order of 5%.

In classical consideration the Shapiro delay is defined by the parameter γ which depends on the theory and in GR $\gamma = 1$. At present the available experimental data are treated such that the best test of γ has been performed in measuring the Shapiro delay when signals from the DSS-25 antenna [70] with the frequencies 7.175GHz and 34.136GHz were sent to the Cassini spacecraft when it was 7AU away from the Earth. The results of the experiment are treated such that $\gamma - 1 = (2.1 \pm 2.3) \cdot 10^{-5}$ [71]. For estimating the quantity $\Delta L(t)$ in this case we take a favorable scenario when the frequency is 34.136GHz and a_{ph} equals the radius of the DSS-25 antenna which is $17m$. Then $\alpha \approx 8 \cdot 10^{-5}$ and $\Delta L(t) \approx 6.7km$ but in view of the WPW effect this quantity will be much greater. This is obviously incompatible with the fact that the accuracy of computing γ is of the order of 10^{-5}.

Our last example is as follows. The astronomical objects called pulsars are treated such that they are neutron stars with radii much less than radii of ordinary stars. Therefore if mechanisms of pulsar electromagnetic radiation were the same as for ordinary stars then the pulsars would not be visible. The fact that pulsars are visible is explained as a consequence of the fact that they emit beams of light which can only be seen when the light is pointed in the direction of the observer with some periods which are treated as periods of rotation of the neutron stars. In popular literature this is compared with the light of a lighthouse. However, by analogy with the case of a signal sent from Cassini, only a small part of photons in the beam can reach the Earth. At present the pulsars have been observed in different regions of the electromagnetic spectrum but the first pulsar called PSR B1919+21 was discovered in 1967 as a radio wave radiation with $\lambda \approx 3.7m$ [72]. This pulsar is treated as the neutron star with the radius $R = 0.97km$ and the distance from the pulsar to Earth is 2283 light years. If for estimating $a_{cl}(t)$ we assume that $a_{cl} = R$ then we get $\alpha \approx 6 \cdot 10^{-4}$ and $a_{cl}(t) \approx 1.3ly \approx 12 \cdot 10^{12}km$. Such an extremely large value of spreading poses a problem whether even predictions of classical electrodynamics are compatible with the fact that pulsars are observable. However, in view of the WPW paradox, the value of $a_{ph}(t)$ will be even much greater and no observation of pulsars would be possible.

Our conclusion is that we have several fundamental paradoxes posing a problem whether predictions of standard quantum theory for the WPS effect are correct.

2.10 Discussion: is it possible to avoid the WPS paradoxes in standard theory?

As shown in the preceding section, if one assumes that photons coming to Earth do not interact with the interstellar or interplanetary medium and with each other then a standard treatment of the WPS effect contradicts the facts that there is no blurring of astronomical images, communication with space probes is possible, the Shapiro delay can be explained in classical theory and GRBs and pulsars are observable. Hence a question arises whether this assumption is legitimate.

As shown in textbooks on quantum optics (see e.g. Ref. [62] and references therein)) quantum states describing the laser emission are strongly coherent and the approximation of independent photons is not legitimate. As shown in Sec. 2.8, the WPS effect in coherent states is pronounced in a much less extent than for individual photons. However, laser emission can be created only at very special conditions when energy levels are inverted, the emission is amplified in the laser cavity etc. At the same time, the main part of the radiation emitted by stars is understood such that it can be approximately described as the blackbody radiation and in addition a part of the radiation consists of photons emitted from different atomic energy levels. In that case the emission of photons is spontaneous rather than induced and one might think that the photons can be treated independently. Several authors (see e.g. Ref. [73] and references therein) discussed a possibility that at some conditions the inverted population and amplification of radiation in stellar atmospheres might occur and so a part of the radiation can be induced. This problem is now under investigation. Hence we adopt a standard assumption that a main part of the radiation from stars is spontaneous. In addition, there is no reason to think that radiation of GRBs, radio antennas, space probes or pulsars is laser like.

The next question is whether the interaction of photons in the above phenomena is important or not. As explained in standard textbooks on QED (see e.g. Ref. [12]), the photon-photon interaction can go only via intermediate creation of virtual electron-positron or quark-antiquark pairs. If ω is the photon frequency, m is the mass of the charged particle in the intermediate state and e is the electric charge of this particle then in the case when $\hbar\omega \ll mc^2$ the total cross section of the photon-photon interaction is [12]

$$\sigma = \frac{56}{5\pi m^2}\frac{139}{90^2}(\frac{e^2}{\hbar c})^4(\frac{\hbar\omega}{mc^2})^6 \tag{2.52}$$

For photons of visible light the quantities $\hbar\omega/(mc^2)$ and σ are very small and for radio waves they are even smaller by several orders of magnitude. At present the effect of the direct photon-photon interaction has not been detected, and experiments with strong laser fields were only able to determine the upper limit of the cross section [74].

The problem of WPS in the ultrarelativistic case has been discussed in a wide literature. As already noted, in Ref. [61] the effect of WPS has been discussed in the Fresnel approximation for a two-dimensional model and the author shows that in the direction perpendicular to the group velocity of the wave spreading is important. He considers WPS in the framework of classical electrodynamics. We believe that considering this effect from quantum point of view is even simpler since the photon wave function satisfies the relativistic Schrödinger equation which is linear in $\partial/\partial t$. As noted in Sec. 2.6, this function also satisfies the wave equation but it is simpler to consider an equation linear in $\partial/\partial t$ than that quadratic in $\partial/\partial t$. However, in classical theory there is no such an object as the photon wave function and hence one has to solve either a system of Maxwell equations or the wave equation. There is also a number of works where the authors consider WPS in view of propagation of classical waves in a medium such that dissipation is important (see e.g. Ref. [75]). In Ref. [76] the effect of WPS has been discussed in view of a possible existence of superluminal neutrinos. The authors consider only the dynamics of the wave packet in the longitudinal direction in the framework of the Dirac equation. They conclude that wave packets describing ultrarelativistic fermions do not experience WPS in this direction. However, the authors do not consider WPS in perpendicular directions.

In view of the above discussion, standard treatment of WPS leads to several fundamental paradoxes of modern theory. To the best of our knowledge, those paradoxes have never been discussed in the literature. For resolving the paradoxes one could discuss several possibilities. One of them might be such that the interaction of light with the interstellar or interplanetary medium cannot be neglected. On quantum level a process of propagation of photons in the medium is rather complicated because several mechanisms of propagation should be taken into account. For example, a possible process is such that a photon can be absorbed by an atom and reemitted. This process makes it clear why the speed of light in the medium is less than c: because the atom which absorbed the photon is in an excited state for some time before reemitting the photon. However, this process is also important from the following point of view: even if the coordinate photon wave function had a large width before absorption, as a consequence of the collapse of the wave function, the wave function of the emitted photon will have in general much smaller dimensions since after detection the width is defined only by parameters of the corresponding detector. If the photon encounters many atoms on its way, this process does not allow the photon wave function to spread out significantly. Analogous remarks can be made about other processes, for example about rescattering of photons on large groups of atoms, rescattering on elementary particles if they are present in the medium etc.

However, such processes have been discussed in Theory B and, as noted in Sec. 2.9, they probably result in more blurring than is seen.

The interaction of photons with the interstellar or interplanetary medium might also be important in view of hypotheses that the density of the medium is much greater than usually believed. Among the most popular scenarios are dark energy, dark matter etc. As shown in our papers (see e.g. Refs. [17, 18, 77]) and in Sec. 3.6), the phenomenon of the cosmological acceleration can be easily and naturally explained from first principles of quantum theory without involving dark energy, empty space-background and other artificial notions. However, the other scenarios seem to be more realistic and one might expect that they will be intensively investigated. A rather hypothetical possibility is that the propagation of photons in the medium has something in common with the induced emission when a photon induces emission of other photons in practically the same direction. In other words, the interstellar medium amplifies the emission as a laser. This possibility seems to be not realistic since it is not clear why the energy levels in the medium might be inverted.

We conclude that at present in standard theory there are no realistic scenarios which can explain the WPS paradoxes. In the remaining part of the chapter we propose a solution of the problem proceeding from a consistent definition of the position operator.

2.11 Consistent construction of position operator

The above results give grounds to think that the reason of the paradoxes which follow from the behavior of the coordinate photon wave function in perpendicular directions is that standard definition of the position operator in those directions does not correspond to realistic measurements of coordinates. Before discussing a consistent construction, let us make the following remark. On elementary level students treat the mass m and the velocity $\mathbf{v}$ as primary quantities such that the momentum is $m\mathbf{v}$ and the kinetic energy is $m\mathbf{v}^2/2$. However, from the point of view of Special Relativity, the primary quantities are the momentum $\mathbf{p}$ and the total energy E and then the mass and velocity are defined as $m^2c^4 = E^2 - \mathbf{p}^2c^2$ and $\mathbf{v} = \mathbf{p}c^2/E$, respectively. This example has the following analogy. In standard quantum theory the primary operators are the position and momentum operators and the orbital angular momentum operator is defined as their cross product. However, the operators $\mathbf{P}$ and $\mathbf{L}$ are consistently defined as representation operators of the Poincare algebra while the definition of the position operator is a problem. Hence a question arises whether the position operator can be defined in terms of $\mathbf{P}$ and $\mathbf{L}$.

One might seek the position operator such that on classical level the relation $\mathbf{r} \times \mathbf{p} = \mathbf{L}$ will take place. Note that on quantum level this relation is not necessary. Indeed, the very fact that some elementary particles have a half-integer spin shows that the total angular momentum for those particles does not have the orbital nature but on classical level the angular momentum can be always represented as a cross product of the radius-vector and standard momentum. However, if the values of $\mathbf{p}$ and $\mathbf{L}$ are known and $\mathbf{p} \neq 0$ then the requirement that $\mathbf{r} \times \mathbf{p} = \mathbf{L}$ does not define $\mathbf{r}$ uniquely. One can define parallel and perpendicular components of $\mathbf{r}$ as $\mathbf{r} = r_{||}\mathbf{p}/p + \mathbf{r}_\perp$ where $p = |\mathbf{p}|$. Then the relation $\mathbf{r} \times \mathbf{p} = \mathbf{L}$ defines uniquely only $\mathbf{r}_\perp$. Namely, as follows from this relation, $\mathbf{r}_\perp = (\mathbf{p} \times \mathbf{L})/p^2$. On quantum level $\mathbf{r}_\perp$ should be replaced by a selfadjoint operator $\mathcal{R}_\perp$ defined as

$$\begin{aligned}\mathcal{R}_{\perp j} &= \frac{\hbar}{2p^2} e_{jkl}(p_k L_l + L_l p_k) = \frac{\hbar}{p^2} e_{jkl} p_k L_l - \frac{i\hbar}{p^2} p_j \\ &= i\hbar \frac{\partial}{\partial p_j} - i\frac{\hbar}{p^2} p_j p_k \frac{\partial}{\partial p_k} - \frac{i\hbar}{p^2} p_j \end{aligned} \tag{2.53}$$

where e_{jkl} is the absolutely antisymmetric tensor, $e_{123} = 1$, a sum over repeated indices is assumed and we assume that if $\mathbf{L}$ is given by Eq. (2.12) then the orbital momentum is $\hbar\mathbf{L}$.

We define the operators $\mathbf{F}$ and $\mathbf{G}$ such that $\mathcal{R}_\perp = \hbar\mathbf{F}/p$ and $\mathbf{G}$ is the operator of multiplication by the unit vector $\mathbf{n} = \mathbf{p}/p$. A direct calculation shows that these operators satisfy the following relations:

$$\begin{aligned}&[L_j, F_k] = ie_{jkl}F_l, \quad [L_j, G_k] = ie_{jkl}F_l, \quad \mathbf{G}^2 = 1, \quad \mathbf{F}^2 = \mathbf{L}^2 + 1 \\ &[G_j, G_k] = 0, \quad [F_j, F_k] = -ie_{jkl}L_l, \quad e_{jkl}\{F_k, G_l\} = 2L_j \\ &\mathbf{LG} = \mathbf{GL} = \mathbf{LF} = \mathbf{FL} = 0, \quad \mathbf{FG} = -\mathbf{GF} = i\end{aligned} \tag{2.54}$$

The first two relations show that $\mathbf{F}$ and $\mathbf{G}$ are the vector operators as expected. The result for the anticommutator shows that on classical level $\mathbf{F} \times \mathbf{G} = \mathbf{L}$ and the last two relations show that on classical level the operators in the triplet $(\mathbf{F}, \mathbf{G}, \mathbf{L})$ are mutually orthogonal.

Note that if the momentum distribution is narrow and such that the mean value of the momentum is directed along the z axis then it does not mean that on the operator level the z component of the operator $\mathcal{R}_\perp$ should be zero. The matter is that the direction of the momentum does not have a definite value. One might expect that only the mean value of the operator $\mathcal{R}_\perp$ will be zero or very small.

In addition, an immediate consequence of the definition (2.53) follows: *Since the momentum and angular momentum operators commute with the Hamiltonian, the distribution of all the components of* $\mathbf{r}_\perp$ *does not depend on time. In particular, there is no WPS in directions defined by* $\mathcal{R}_\perp$. This is also clear from the fact that $\mathcal{R}_\perp = \hbar\mathbf{F}/p$ where the operator $\mathbf{F}$ acts only over angular variables and the Hamiltonian depends only on p. On classical level the conservation of $\mathcal{R}_\perp$ is obvious since it is defined by the conserving quantities $\mathbf{p}$ and $\mathbf{L}$. It is also obvious that since a free particle is moving along a straight line, a vector from the origin perpendicular to this line does not change with time.

The above definition of the perpendicular component of the position operator is well substantiated since on classical level the relation $\mathbf{r} \times \mathbf{p} = \mathbf{L}$ has been verified in numerous experiments. However, this relation does not make it possible to define the parallel component of the position operator and a problem arises what physical arguments should be used for that purpose.

A direct calculation shows that if $\partial/\partial\mathbf{p}$ is written in terms of p and angular variables then

$$i\hbar\frac{\partial}{\partial\mathbf{p}} = \mathbf{G}\mathcal{R}_{||} + \mathcal{R}_{\perp} \tag{2.55}$$

where the operator $\mathcal{R}_{||}$ acts only over the variable p:

$$\mathcal{R}_{||} = i\hbar(\frac{\partial}{\partial p} + \frac{1}{p}) \tag{2.56}$$

The correction $1/p$ is related to the fact that the operator $\mathcal{R}_{||}$ is Hermitian since in variables $(p, \mathbf{n})$ the scalar product is given by

$$(\chi_2, \chi_1) = \int \chi_2(p, \mathbf{n})^* \chi_1(p, \mathbf{n}) p^2 dp do \tag{2.57}$$

where do is the element of the solid angle.

While the components of standard position operator commute with each other, the operators $\mathcal{R}_{||}$ and $\mathcal{R}_{\perp}$ satisfy the following commutation relation:

$$[\mathcal{R}_{||}, \mathcal{R}_{\perp}] = -\frac{i\hbar}{p}\mathcal{R}_{\perp}, \quad [\mathcal{R}_{\perp j}, \mathcal{R}_{\perp k}] = -\frac{i\hbar^2}{p^2}e_{jkl}L_l \tag{2.58}$$

An immediate consequence of these relation follows: *Since the operator $\mathcal{R}_{||}$ and different components of $\mathcal{R}_{\perp}$ do not commute with each other, the corresponding quantities cannot be simultaneously measured and hence there is no wave function $\psi(r_{||}, \mathbf{r}_{\perp})$ in coordinate representation.*

In standard theory $-\hbar^2(\partial/\partial\mathbf{p})^2$ is the operator of the quantity $\mathbf{r}^2$. As follows from Eq. (2.54), the two terms in Eq. (2.55) are not strictly orthogonal and on the operator level $-\hbar^2(\partial/\partial\mathbf{p})^2 \neq \mathcal{R}_{||}^2 + \mathcal{R}_{\perp}^2$. A direct calculation using Eqs. (2.54) and (2.55) gives

$$\frac{\partial^2}{\partial\mathbf{p}^2} = \frac{\partial^2}{\partial p^2} + \frac{2}{p}\frac{\partial}{\partial p} - \frac{\mathbf{L}^2}{p^2}, \quad -\hbar^2\frac{\partial^2}{\partial\mathbf{p}^2} = \mathcal{R}_{||}^2 + \mathcal{R}_{\perp}^2 - \frac{\hbar^2}{p^2} \tag{2.59}$$

in agreement with the expression for the Laplacian in spherical coordinates. In semiclassical approximation, $(\hbar^2/p^2) \ll \mathcal{R}_{\perp}^2$ since the eigenvalues of $\mathbf{L}^2$ are $l(l+1)$, in semiclassical states $l \gg 1$ and, as follows from Eq. (2.54), $\mathcal{R}_{\perp}^2 = [\hbar^2(l^2+l+1)/p^2]$.

As follows from Eq. (2.58), $[\mathcal{R}_{||}, p] = -i\hbar$, i.e. in the longitudinal direction the commutation relation between the coordinate and momentum is the same as in standard theory. One can also calculate the commutators between the different components of $\mathcal{R}_{\perp}$ and $\mathbf{p}$. Those commutators are not given by such simple expressions as in standard theory but it is easy to see that all of them are of the order of $\hbar$ as it should be.

Equation (2.55) can be treated as an implementation of the relation $\mathbf{r} = r_{||}\mathbf{p}/|\mathbf{p}| + \mathbf{r}_\perp$ on quantum level. As argued in Secs. 2.1 and 2.2, the standard position operator $i\hbar\partial/\partial p_j$ in the direction j is not consistently defined if p_j is not sufficiently large. One might think however that since the operator $\mathcal{R}_{||}$ contains $i\hbar\partial/\partial p$, it is defined consistently if the magnitude of the momentum is sufficiently large.

In summary, we propose to define the position operator not by the set $(i\hbar\partial/\partial p_x, i\hbar\partial/\partial p_y, i\hbar\partial/\partial p_z)$ but by the operators $\mathcal{R}_{||}$ and $\mathcal{R}_\perp$. Those operators are defined from different considerations. As noted above, the definition of $\mathcal{R}_\perp$ is based on solid physical facts while the definition of $\mathcal{R}_{||}$ is expected to be more consistent than the definition of standard position operator. However, this does not guarantee that the operator $\mathcal{R}_{||}$ is consistently defined in all situations. As argued in Sec. 5.3, in a quantum theory over a Galois field an analogous definition is not consistent *for macroscopic bodies* (even if p is large) since in that case semiclassical approximation is not valid. In the remaining part of the chapter we assume that for elementary particles the above definition of $\mathcal{R}_{||}$ is consistent in situations when semiclassical approximation applies.

One might pose the following question. What is the reason to work with the parallel and perpendicular components of the position operator separately if, according to Eq. (2.55), their sum is the standard position operator? The explanation follows.

In quantum theory every physical quantity corresponds to a selfadjoint operator but the theory does not define explicitly how a quantity corresponding to a specific operator should be measured. There is no guaranty that for each selfadjoint operator there exists a physical quantity which can be measured in real experiments.

Suppose that there are three physical quantities corresponding to the selfadjoint operators A, B and C such that $A + B = C$. Then in each state the mean values of the operators are related as $\bar{A} + \bar{B} = \bar{C}$ but in situations when the operators A and B do not commute with each other there is no direct relation between the distributions of the physical quantities corresponding to the operators A, B and C. For example, in situations when the physical quantities corresponding to the operators A and B are semiclassical and can be measured with a good accuracy, there is no guaranty that the physical quantity corresponding to the operator C can be measured in real measurements. As an example, the physical meaning of the quantity corresponding to the operator $L_x + L_y$ is problematic. Another example is the situation with WPS in directions perpendicular to the particle momentum. Indeed, as noted above, the physical quantity corresponding to the operator $\mathcal{R}_\perp$ does not experience WPS and, as shown in Sec. 2.13, in the case of ultrarelativistic particles there is no WPS in the parallel direction as well. However, standard position operator is a sum of noncommuting operators corresponding to well defined physical quantities and, as a consequence, there are situations when standard position operator defines a quantity which cannot be measured in real experiments.

2.12 New position operator and semiclassical states

As noted in Sec. 2.2, in standard theory states are treated as semiclassical in greatest possible extent if $\Delta r_j \Delta p_j = \hbar/2$ for each j and such states are called coherent. The existence of coherent states in standard theory is a consequence of commutation relations $[p_j, r_k] = -i\hbar\delta_{jk}$. Since in our approach there are no such relations, a problem arises how to construct states in which all physical quantities p, $r_{||}$, $\mathbf{n}$ and $\mathbf{r}_\perp$ are semiclassical.

One of the ways to prove this is to calculate the mean values and uncertainties of the operator $\mathcal{R}_{||}$ and all the components of the operator $\mathcal{R}_\perp$ in the state defined by Eq. (2.17). The calculation is not simple since it involves three-dimensional integrals with Gaussian functions divided by p^2. The result is that these operators are semiclassical in the state (2.17) if $p_0 \gg \hbar/b$, $p_0 \gg \hbar/a$ and r_{0z} has the same order of magnitude as r_{0x} and r_{0y}.

However, a more natural approach is as follows. Since $\mathcal{R}_\perp = \hbar\mathbf{F}/p$, the operator $\mathbf{F}$ acts only over the angular variable $\mathbf{n}$ and $\mathcal{R}_{||}$ acts only over the variable p, it is convenient to work in the representation where the Hilbert space is the space of functions $\chi(p, l, \mu)$ such that the scalar product is

$$(\chi_2, \chi_1) = \sum_{l\mu} \int_0^\infty \chi_2(p, l, \mu)^* \chi_1(p, l, \mu) dp \tag{2.60}$$

and l and μ are the orbital and magnetic quantum numbers, respectively, i.e.

$$\mathbf{L}^2\chi(p, l, \mu) = l(l + 1)\chi(p, l, \mu), \quad L_z\chi(p, l, \mu) = \mu\chi(p, l, \mu) \tag{2.61}$$

The operator $\mathbf{L}$ in this space does not act over the variable p and the action of the remaining components is given by

$$L_+\chi(l, \mu) = [(l+\mu)(l+1-\mu)]^{1/2}\chi(l, \mu-1), \quad L_-\chi(l, \mu) = [(l-\mu)(l+1+\mu)]^{1/2}\chi(l, \mu+1) \tag{2.62}$$

where the $\pm$ components of vectors are defined such that $L_x = L_+ + L_-$, $L_y = -i(L_+ - L_-)$.

A direct calculation shows that, as a consequence of Eq. (2.53)

$$F_+\chi(l,\mu) = -\frac{i}{2}[\frac{(l+\mu)(l+\mu-1)}{(2l-1)(2l+1)}]^{1/2} l\chi(l-1,\mu-1)$$
$$-\frac{i}{2}[\frac{(l+2-\mu)(l+1-\mu)}{(2l+1)(2l+3)}]^{1/2}(l+1)\chi(l+1,\mu-1)$$
$$F_-\chi(l,\mu) = \frac{i}{2}[\frac{(l-\mu)(l-\mu-1)}{(2l-1)(2l+1)}]^{1/2} l\chi(l-1,\mu+1)$$
$$+\frac{i}{2}[\frac{(l+2+\mu)(l+1+\mu)}{(2l+1)(2l+3)}]^{1/2}(l+1)\chi(l+1,\mu+1)$$
$$F_z\chi(l,\mu) = i[\frac{(l-\mu)(l+\mu)}{(2l-1)(2l+1)}]^{1/2} l\chi(l-1,\mu)$$
$$-i[\frac{(l+1-\mu)(l+1+\mu)}{(2l+1)(2l+3)}]^{1/2}(l+1)\chi(l+1,\mu) \qquad (2.63)$$

The operator **G** acts on such states as follows

$$G_+\chi(l,\mu) = \frac{1}{2}[\frac{(l+\mu)(l+\mu-1)}{(2l-1)(2l+1)}]^{1/2}\chi(l-1,\mu-1)$$
$$-\frac{1}{2}[\frac{(l+2-\mu)(l+1-\mu)}{(2l+1)(2l+3)}]^{1/2}\chi(l+1,\mu-1)$$
$$G_-\chi(l,\mu) = -\frac{1}{2}[\frac{(l-\mu)(l-\mu-1)}{(2l-1)(2l+1)}]^{1/2}\chi(l-1,\mu+1)$$
$$+\frac{1}{2}[\frac{(l+2+\mu)(l+1+\mu)}{(2l+1)(2l+3)}]^{1/2}\chi(l+1,\mu+1)$$
$$G_z\chi(l,\mu) = -[\frac{(l-\mu)(l+\mu)}{(2l-1)(2l+1)}]^{1/2}\chi(l-1,\mu)$$
$$-[\frac{(l+1-\mu)(l+1+\mu)}{(2l+1)(2l+3)}]^{1/2}\chi(l+1,\mu) \qquad (2.64)$$

and now the operator $\mathcal{R}_{||}$ has a familiar form $\mathcal{R}_{||} = i\hbar\partial/\partial p$.

Therefore by analogy with Secs. 2.2 and 2.3 one can construct states which are coherent with respect to $(r_{||}, p)$, i.e. such that $\Delta r_{||}\Delta p = \hbar/2$. Indeed (see Eq. (2.4)), the wave function

$$\chi(p) = \frac{b^{1/2}}{\pi^{1/4}\hbar^{1/2}} exp[-\frac{(p-p_0)^2 b^2}{2\hbar^2} - \frac{i}{\hbar}(p-p_0)r_0] \qquad (2.65)$$

describes a state where the mean values of p and $r_{||}$ are p_0 and r_0, respectively and their uncertainties are $\hbar/(b\sqrt{2})$ and $b/\sqrt{2}$, respectively. Strictly speaking, the analogy between the given case and that discussed in Secs. 2.2 and 2.3 is not full since in the given case the quantity p can be in the range $[0.\infty)$, not in $(-\infty,\infty)$ as momentum variables used in those sections. However, if $p_0 b/\hbar \gg 1$ then the formal expression for $\chi(p)$ at $p<0$ is extremely small and so the normalization integral for $\chi(p)$ can be formally taken from $-\infty$ to ∞.

In such an approximation one can define wave functions $\psi(r)$ in the $r_{||}$ representation. By analogy with the consideration in Secs. 2.2 and 2.3 we define

$$\psi(r) = \int exp(\frac{i}{\hbar}pr)\chi(p)\frac{dp}{(2\pi\hbar)^{1/2}} \tag{2.66}$$

where the integral is formally taken from $-\infty$ to ∞. Then

$$\psi(r) = \frac{1}{\pi^{1/4}b^{1/2}}exp[-\frac{(r-r_0)^2}{2b^2}+\frac{i}{\hbar}p_0 r] \tag{2.67}$$

Note that here the quantities r and r_0 have the meaning of coordinates in the direction parallel to the particle momentum, i.e. they can be positive or negative.

Consider now states where the quantities $\mathbf{F}$ and $\mathbf{G}$ are semiclassical. One might expect that in semiclassical states the quantities l and μ are very large. In this approximation, as follows from Eqs. (2.63) and (2.64), the action of the operators $\mathbf{F}$ and $\mathbf{G}$ can be written as

$$\begin{aligned}
F_+\chi(l,\mu) &= -\frac{i}{4}(l+\mu)\chi(l-1,\mu-1) - \frac{i}{4}(l-\mu)\chi(l+1,\mu-1)\\
F_-\chi(l,\mu) &= \frac{i}{4}(l-\mu)\chi(l-1,\mu+1) + \frac{i}{4}(l+\mu)\chi(l+1,\mu+1)\\
F_z\chi(l,\mu) &= -\frac{i}{2l}(l^2-\mu^2)^{1/2}[\chi(l+1,\mu)-\chi(l-1,\mu)]\\
G_+\chi(l,\mu) &= \frac{l+\mu}{4l}\chi(l-1,\mu-1) - \frac{l-\mu}{4l}\chi(l+1,\mu-1)\\
G_-\chi(l,\mu) &= -\frac{l-\mu}{4l}\chi(l-1,\mu+1) + \frac{l+\mu}{4l}\chi(l+1,\mu+1)\\
G_z\chi(l,\mu) &= -\frac{1}{2l}(l^2-\mu^2)^{1/2}[\chi(l+1,\mu)+\chi(l-1,\mu)]
\end{aligned} \tag{2.68}$$

In view of the remark in Sec. 2.2 about semiclassical vector quantities, consider a state $\chi(l,\mu)$ such that $\chi(l,\mu)\neq 0$ only if $l\in[l_1,l_2]$, $\mu\in[\mu_1,\mu_2]$ where $l_1,\mu_1>0$, $\delta_1 = l_2+1-l_1$, $\delta_2=\mu_2+1-\mu_1$, $\delta_1 \ll l_1$, $\delta_2 \ll \mu_1$ $\mu_2<l_1$ and $\mu_1 \gg (l_1-\mu_1)$. This is the state where the quantity μ is close to its maximum value l. As follows from Eqs. (2.61) and (2.62), in this state the quantity L_z is much greater than L_x and L_y and, as follows from Eq. (2.68), the quantities F_z and G_z are small. So on classical level this state describes a motion of the particle in the xy plane. The quantity L_z in this state is obviously semiclassical since $\chi(l,\mu)$ is the eigenvector of the operator L_z with the eigenvalue μ. As follows from Eq. (2.68), the action of the operators (F_+,F_-,G_+,G_-) on this state can be described by the

following approximate formulas:

$$F_+\chi(l,\mu) = -\frac{il_0}{2}\chi(l-1,\mu-1), \quad F_-\chi(l,\mu) = \frac{il_0}{2}\chi(l+1,\mu+1)$$
$$G_+\chi(l,\mu) = \frac{1}{2}\chi(l-1,\mu-1), \quad G_-\chi(l,\mu) = \frac{1}{2}\chi(l+1,\mu+1) \tag{2.69}$$

where l_0 is a value from the interval $[l_1, l_2]$.

Consider a simple model when $\chi(l,\mu) = exp[i(l\alpha - \mu\beta)]/(\delta_1\delta_2)^{1/2}$ when $l \in [l_1, l_2]$ and $\mu \in [\mu_1, \mu_2]$. Then a simple direct calculation using Eq. (2.69) gives

$$\bar{G}_x = cos\gamma, \quad \bar{G}_y = -sin\gamma, \quad \bar{F}_x = -l_0 sin\gamma, \quad \bar{F}_y = -l_0 cos\gamma$$
$$\Delta G_x = \Delta G_y = (\frac{1}{\delta_1} + \frac{1}{\delta_2})^{1/2}, \quad \Delta F_x = \Delta F_y = l_0(\frac{1}{\delta_1} + \frac{1}{\delta_2})^{1/2} \tag{2.70}$$

where $\gamma = \alpha - \beta$. Hence the vector quantities **F** and **G** are semiclassical since either $|cos\gamma|$ or $|sin\gamma|$ or both are much greater than $(\delta_1 + \delta_2)/(\delta_1\delta_2)$.

2.13 New position operator and WPS

If the space of states is implemented according to the scalar product (2.60) then the dependence of the wave function on t is

$$\chi(p,k,\mu,t) = exp[-\frac{i}{\hbar}(m^2c^2 + p^2)^{1/2}ct]\chi(p,k,\mu,t=0) \tag{2.71}$$

As noted in Secs. 2.3 and 2.5, there is no WPS in momentum space and this is natural in view of momentum conservation. Then, as already noted, the distribution of the quantity $\mathbf{r}_\perp$ does not depend on time and this is natural from the considerations described in Sec. 2.11.

At the same time, the dependence of the $r_{||}$ distribution on time can be calculated in full analogy with Sec. 2.3. Indeed, consider, for example a function $\chi(p,l,\mu,t=0)$ having the form

$$\chi(p,l,\mu,t=0) = \chi(p,t=0)\chi(l,\mu) \tag{2.72}$$

Then, as follows from Eqs. (2.66) and (2.71),

$$\psi(r,t) = \int exp[-\frac{i}{\hbar}(m^2c^2 + p^2)^{1/2}ct + \frac{i}{\hbar}pr]\chi(p,t=0)\frac{dp}{(2\pi\hbar)^{1/2}} \tag{2.73}$$

Suppose that the function $\chi(p,t=0)$ is given by Eq. (2.65). Then in full analogy with the calculations in Sec. 2.3 we get that in the nonrelativistic case the $r_{||}$ distribution is defined by the wave function

$$\psi(r,t) = \frac{1}{\pi^{1/4}b^{1/2}}(1+\frac{i\hbar t}{mb^2})^{-1/2}exp[-\frac{(r-r_0-v_0t)^2}{2b^2(1+\frac{\hbar^2t^2}{m^2b^4})}(1-\frac{i\hbar t}{mb^2})+\frac{i}{\hbar}p_0r-\frac{ip_0^2t}{2m\hbar}] \tag{2.74}$$

where $v_0 = p_0/m$ is the classical speed of the particle in the direction of the particle momentum. Hence the WPS effect in this direction is similar to that given by Eq. (2.8) in standard theory.

In the opposite case when the particle is ultrarelativistic, Eq. (2.73) can be written as

$$\psi(r,t) = \int exp[\frac{i}{\hbar}p(r-ct)]\chi(p,t=0)\frac{dp}{(2\pi\hbar)^{1/2}} \tag{2.75}$$

Hence, as follows from Eq. (2.67):

$$\psi(r,t) = \frac{1}{\pi^{1/4}b^{1/2}}exp[-\frac{(r-r_0-ct)^2}{2b^2} + \frac{i}{\hbar}p_0(r-ct)] \tag{2.76}$$

In particular, for an ultrarelativistic particle there is no WPS in the direction of particle momentum and this is in agreement with the results of Sec. 2.5.

We conclude that in our approach an ultrarelativistic particle (e.g. the photon) experiences WPS neither in the direction of its momentum nor in transverse directions, i.e. the WPS effect for an ultrarelativistic particle is absent at all.

Let us note that the absence of WPS in transverse directions is simply a consequence of the fact that a consistently defined operator $\mathcal{R}_\perp$ commutes with the Hamiltonian, i.e. $\mathbf{r}_\perp$ is a conserving physical quantity. On the other hand, the longitudinal coordinate cannot be conserving since a particle is moving along the direction of its momentum. However, in a special case of ultrarelativistic particle the absence of WPS is simply a consequence of the fact that the wave function given by Eq. (2.75) depends on r and t only via a combination of $r - ct$.

Consider now the spherically symmetric model discussed in Sec. 2.9 when the momentum wave function is described by Eq. (2.42). As noted in Sec. 2.9, this state is not semiclassical and hence a choice of the position operator adequately describing this state is problematic. As noted in Sec. 2.9, the standard choice leads to the result given by Eq. (2.51) which is counterintuitive. We now consider what happens when the state (2.42) is described by the position operator proposed in Sec. 2.11.

Since the angular momentum of the state (2.42) is zero, it follows from Eq. (2.53) that $||\mathcal{R}_\perp\chi|| \leq \lambda||\chi||$. Therefore when we consider large distances, the contribution of $\mathcal{R}_\perp$ can be neglected. Then, as follows from Eq. (2.55), the position operator in this situation is $\mathcal{R} = \mathbf{G}\mathcal{R}_{||}$. In this approximation different components of the position operator commute with each other. Therefore one can define the coordinate wave function which in the given case again has the form (2.46).

Since $\mathbf{p} = \mathbf{G}p$, $\mathbf{G}$ acts only on angular variables and $\mathcal{R}_{||}$ acts only on the variable p we conclude that in the given case the angular parts of the position and momentum operators are the same in contrast to the situation in standard theory where those parts are related to each other by the Fourier transform.

As noted in Sec. 2.9, in standard theory the angular resolution corresponding to Eq. (2.51) is a quantity of the order of λ/d while from obvious geometrical considerations one might think that this quantity should be of the order of R/L. However, now the value of the angular resolution is exactly what it should be from the geomentical considerations.

Indeed, as noted in Sec. 2.9, for calculating the angular resolution one should project the coordinate wave function on the state having the support inside the volume V where the photon will be measured. Suppose that the coordinate wave function is spherically symmetric with respect to the origin characterized by $r_0 = 0$ and one of the conditions characterizing the volume is such that the angular variables are in the range within the element do of the solid angle. Then in view of the fact that angular variables in the coordinate and momentum wave functions are the same, any measurement of the photon momentum inside V can give only the results where the photon momentum is inside do. Therefore, as noted in Sec. 2.9, for a pointlike source of light the angular resolution is of the order of d/L and for a star with the radius R the resolution is of the order of R/L. Hence, in contrast to the situation discussed in Sec. 2.9, there is no blurring of astronomical images because the angular resolution is always ideal and does not depend on d. However, details of astronomical objects will be distinguishable only if d is rather large because, as follows from Eq. (2.49), the norm of the function $\mathcal{P}\psi(r,t)$ is of the order of d/L.

2.14 Summary

In this chapter we consider a problem of constructing position operator in quantum theory. As noted in Sec. 2.1, this operator is needed in situations where semiclassical approximation works with a high accuracy and the example with the spherically symmetric case discussed at the end of the preceding section indicates that this operator can be useful in other problems.

A standard choice of the position operator in momentum space is $i\hbar\partial/\partial\mathbf{p}$. A motivation for this choice is discussed in Sec. 2.2. We note that the standard definition is not consistent since $i\hbar\partial/\partial p_j$ cannot be a physical position operator in directions where the momentum is small. Physicists did not pay attention to the inconsistency probably for the following reason: as explained in standard textbooks on quantum mechanics, the transition from quantum to classical theory can be performed such that if the coordinate wave function contains a rapidly oscillating exponent $exp(iS/\hbar)$ where S is the classical action then in the formal limit $\hbar \to 0$ the Schrödinger equation becomes the Hamilton-Jacobi equation.

However, an inevitable consequence of standard quantum theory is the effect of wave packet spreading (WPS). This fact has not been considered as a drawback of the theory. Probably the reasons are that for macroscopic bodies this effect is extremely small while in experiments on the Earth with atoms and elementary particles spreading probably does not have enough time to manifest itself. However, for photons travelling to the Earth from distant objects this effect is considerable, and it seems that this fact has been overlooked by physicists.

As shown in Sec. 2.9, if the WPS effect for photons travelling to Earth from distant objects is as given by standard theory then we have several fundamental paradoxes: a) if the major part of photons emitted by stars are in wave packet states (what is the most probable scenario) then we should see not stars but only an almost continuous background from all stars; b) no anisotropy of the relic radiation could be observable; c) the effect of WPS is incompatible with the data on gamma-ray bursts; d) communication with distant space probes could not be possible; e) the Shapiro delay could not be explained only in the framework of classical theory; f) the fact that we can observe pulsars could not be explained. In addition, the consideration in Secs. 2.9 and 2.13 poses the following questions: g) how is it possible to verify that the angular resolution of a star in the part of the spectrum corresponding to transitions between atomic levels is of the order of λ/d rather than R/L?; h) are predictions of standard theory on the WPS effect for protons in the LHC ring compatible with experimental data? We have also noted that in the scenario when the quantities $N_\perp$ are not very large, even images of planets will be blurred.

In Sec. 2.7 it is shown that, from the point of view of standard quantum theory, there exists the WPW paradox that after some period of time the transversal widths of the coordinate wave functions for photons comprising a classical wave packet will be typically much greater than the transversal width of the classical packet as a whole. This situation seems to be fully unphysical since, as noted in Sec. 2.7, different photons in a classical wave packet do not interfere with each other. The calculations in Sec. 2.5 show that the reason of the WPW paradox is that in directions perpendicular to the particle momentum the standard position operator is defined inconsistently. At the same time, as shown in Sec. 2.8, fow coherent states the WPS effect is pronounced in a much less extent than for individual photons.

We propose a new definition of the position operator which we treat as consistent for the following reasons. Our position operator is defined by two components - in the direction along the momentum and in perpendicular directions. The first part has a familiar form $i\hbar\partial/\partial p$ and is treated as the operator of the longitudinal coordinate if the magnitude of p is rather large. At the same condition the position operator in the perpendicular directions is defined as a quantum generalization of the relation $\mathbf{r}_\perp \times \mathbf{p} = \mathbf{L}$. So in contrast to the standard definition of the position operator, the new operator is expected to be physical only if the *magnitude* of the momentum is rather large.

As a consequence of our construction, WPS in directions perpendicular to the particle momentum is absent regardless of whether the particle is nonrelativistic or relativistic. Moreover, for an ultrarelativistic particle the effect of WPS is absent at all.

As noted in Sec. 2.7, in standard quantum theory photons comprising a classical electromagnetic wave packet cannot be (approximately) treated as point-like particles in view of the WPW paradox. However, in our approach, in view of the absence of WPS for massless particles, the usual intuition is restored and photons comprising a divergent classical wave packet can be (approximately) treated as pointlike particles. Moreover, the phenomenon of divergence of a classical wave packet can now be naturally explained simply as a consequence of the fact that different photons in the packet have different momenta.

Our result resolves the above paradoxes and, in view of the above discussion, also poses a problem whether the results of classical electrodynamics can be applied for wave packets moving for a long period of time. For example, as noted in Sec. 2.9, even classical theory predicts that when a wave packet emitted in gamma-ray bursts or by a pulsar reaches the Earth, the width of the packet is extremely large (while the value predicted by standard quantum theory is even much greater) and this poses a problem whether such a packet can be detected. We believe that a natural explanation of why classical theory does not apply in this case is as follows. As noted in Sec. 2.4, classical electromagnetic fields should be understood as a result of taking mean characteristics for many photons. Then the fields will be (approximately) continuous if the density of the photons is high. However, for a divergent beam of photons their density decreases with time. Hence after a long period of time the mean characteristics of the photons in the beam cannot represent continuous fields. In other words, in this situation the set of photons cannot be effectively described by classical electromagnetic fields.

A picture that a classical wave packet can be treated as a collection of (almost) pointlike photons also sheds new light on the explanation of known phenomena. Suppose that a wide beam of visible light falls on a screen which is perpendicular to the direction of light. Suppose that the total area of the screen is S but the surface contains slits with the total area S_1. We are interested in the question of what part of the light will pass the screen. The answer that the part equals S_1/S follows from the picture that the light consists of many almost pointlike photons moving along geometrical trajectories and hence only the S_1/S part of the photons will pass the surface. Numerous experiments show that deviations from the above answer begin to manifest in interference experiments where dimensions of slits and distances between them have the order of tens or hundreds of microns or even less. In classical theory interference is explained as a phenomenon arising when the wave length of the classical electromagnetic wave becomes comparable to dimensions of slits and distances between them. However, as noted in Sec. 2.4, the notion of wave length does not have the usual meaning on quantum level. From the point of view of particle theory, the phenomenon of interference has a natural explanation that it occurs when dimensions of slits and distances between them become comparable to the typical width of the photon wave function.

Our results on the position operator also pose a problem how the interference phenomenon should be explained on the level of single photons. The usual qualitative explanation is as follows. Suppose that the mean momentum of a photon is directed along the z axis perpendicular to a screen. If the (x, y) dependence of the photon wave function is highly homogeneous then the quantities Δp_x and Δp_y

are very small. When the photon passes the screen with holes, its wave function is not homogeneous in the xy plane anymore. As a result, the quantities Δp_x and Δp_y become much greater and the photon can be detected in points belonging to the geometrical shadow. However, such an explanation is problematic for the following reason. Since the mean values of the x and y components of the photon momentum are zero, as noted in Secs. 2.2 and 2.4, the (x, y) dependence of the wave function cannot be semiclassical and, as it has been noted throughout the paper, in that case standard position operator in the xy plane is not consistently defined.

Different components of the new position operator do not commute with each other and, as a consequence, there is no wave function in coordinate representation. In particular, there is no quantum analog of the coordinate Coulomb potential (see the discussion in Sec. 2.1). A possibility that coordinates can be noncommutative has been first discussed by Snyder [78] and it is implemented in several modern theories. In those theories the measure of noncommutativity is defined by a parameter l called the fundamental length (the role of which can be played e.g. by the Planck length or the Schwarzschild radius). In the formal limit $l \to 0$ the coordinates become standard ones related to momenta by a Fourier transform. As shown in Sec. 2.9, this is unacceptable in view of the WPS paradoxes. One of ideas of those theories is that with a nonzero l it might be possible to resolve difficulties of standard theory where $l = 0$ (see e.g. Ref. [79] and references therein). At the same time, in our approach there can be no notion of fundamental length since commutativity of coordinates takes place only in the formal limit $\hbar \to 0$.

The position operator proposed in this chapter might be also important in view of the following. There exists a wide literature discussing the Einstein-Podolsky-Rosen paradox, locality in quantum theory, quantum entanglement, Bell's theorem and similar problems (see e.g. Ref. [58] and references therein). Consider, for example, the following problem in standard theory. Let at $t = 0$ particles 1 and 2 be localized inside finite volumes V_1 and V_2, respectively, such that the volumes are very far from each other. Hence the particles don't interact with each other. However, as follows from Eq. (2.14), their wave functions will overlap at any $t > 0$ and hence the interaction can be transmitted even with an infinite speed. This is often characterized as quantum nonlocality, entanglement and/or action at a distance.

Consider now this problem in the framework of our approach. Since in this approach there is no wave function in coordinate representation, there is no notion of a particle localized inside a finite volume. In addition, as noted in Sec. 1.2, standard treatment of time might be problematic. Hence a problem arises whether on quantum level the notions of locality or nonlocality have a physical meaning. In our approach spreading does not take place in directions perpendicular to the particle momenta and for ultrarelativistic particles spreading does not occur at all. Hence, at least in the case of ultrarelativistic particles, this kind of interaction does not occur in agreement with classical intuition that no interaction can be transmitted with the speed greater than c. This example poses a problem whether the position operator should be modified not only in directions perpendicular to particle momenta but also in longitudinal directions such that the effect of WPS should be excluded at all.

The above discussion shows that the problem of transition from quantum theory to classical one should be reformulated. This is not an academic but extremely important problem of modern physics. Indeed, if we believe that quantum theory is fundamental then it should describe not only atoms and elementary particles but even the motion of bodies in the Solar System and in the World. So we need to know how the evolution of macroscopic bodies should be described in quantum theory and what is the correct choice of position operator.

As noted above, in directions perpendicular to the particle momentum the choice of the position operation is based only on the requirement that semiclassical approximation should reproduce the standard relation $\mathbf{r}_\perp \times \mathbf{p} = \mathbf{L}$. This requirement seems to be beyond any doubts since *on classical level* this relation is confirmed in numerous experiments. At the same time, the choice $i\hbar\partial/\partial p$ of the coordinate operator in the longitudinal direction is analogous to that in standard theory and hence one might expect that this operator is physical if the magnitude of p is rather large.

As shown in Chap. 4, the construction of the position operator described in this chapter for the case of Poincare invariant theory can be generalized to the case of de Sitter (dS) invariant theory. In this case the interpretation of the position operator is even more important than in Poincare invariant theory. The reason is that even the free two-body mass operator in the dS theory depends not only on the relative two-body momentum but also on the distance between the particles.

As argued in Chap. 5, in dS theory over a Galois field the assumption that the dS analog of the operator $i\hbar\partial/\partial p$ is the operator of the longitudinal coordinate is not valid *for macroscopic bodies* (even if p is large) since in that case semiclassical approximation is not valid. We have proposed a modification of the position operator such that quantum theory reproduces for the two-body mass operator the mean value compatible with the Newton law of gravity and precession of Mercury's perihelion. Then a problem arises how quantum theory can reproduce classical evolution for macroscopic bodies.

Our result for ultrarelativistic particles can be treated as ideal: quantum theory reproduces the motion along a classical trajectory without any spreading. However, this is only a special case of one free elementary particle. If quantum theory is treated as more general than the classical one then it should describe not only elementary particles and atoms but even the motion of macroscopic bodies in the Solar System and in the World. We believe that the assumption that the evolution of macroscopic bodies can be described by the Schrödinger equation is unphysical. For example, if the motion of the Earth is described by the evolution operator $exp[-iH(t_2-t_1)/\hbar]$ where H is the Hamiltonian of the Earth then the quantity $H(t_2-t_1)/\hbar$ becomes of the order of unity when t_2-t_1 is a quantity of the order of $10^{-68}s$ if the Hamiltonian is written in nonrelativistic form and $10^{-76}s$ if it is written in relativistic form. Such time intervals seem to be unphysical and so in the given case the approximation when t is a continuous parameter seems to be unphysical too. In modern theories (e.g. in the Big Bang hypothesis) it is often stated that the Planck time $t_P \approx 10^{-43}s$ is a physical minimum time interval. However, at present there are no experiments confirming that time intervals of the order of $10^{-43}s$ can be measured.

The time dependent Schrödinger equation has not been experimentally verified and the major theoretical arguments in favor of this equation are as follows: a) the Hamiltonian is the generator of the time translation in the Minkowski space; b) this equation becomes the Hamilton-Jacobi one in the formal limit $\hbar \to 0$. However, as argued in Chap. 1, quantum theory should not be based on the space-time background and the conclusion b) is made without taking into account the WPS effect. Hence the problem of describing evolution in quantum theory remains open.

The above examples show that at macroscopic level a consistent definition of the transition from quantum to classical theory is the fundamental open problem.

Chapter 3

Basic properties of dS quantum theories

3.1 dS invariance vs. AdS and Poincare invariance

As already mentioned, one of the motivations for this work is to investigate whether standard gravity can be obtained in the framework of a free theory. In standard nonrelativistic approximation, gravity is characterized by the term $-Gm_1m_2/r$ in the mean value of the mass operator. Here G is the gravitational constant, m_1 and m_2 are the particle masses and r is the distance between the particles. Since the kinetic energy is always positive, the free nonrelativistic mass operator is positive definite and therefore there is no way to obtain gravity in the framework of the free theory. Analogously, in Poincare invariant theory the spectrum of the free two-body mass operator belongs to the interval $[m_1 + m_2, \infty)$ while the existence of gravity necessarily requires that the spectrum should contain values less than $m_1 + m_2$.

In theories where the symmetry algebra is the AdS algebra so(2,3), the structure of IRs is well-known (see e.g. Ref. [80] and Chap. 8). In particular, for positive energy IRs the AdS Hamiltonian has the spectrum in the interval $[m, \infty)$ and m has the meaning of the mass. Therefore the situation is pretty much analogous to that in Poincare invariant theories. In particular, the free two-body mass operator again has the spectrum in the interval $[m_1 + m_2, \infty)$ and therefore there is no way to reproduce gravitational effects in the free AdS invariant theory.

As noted in Sec. 1.4, the existing experimental data practically exclude the possibility that $\Lambda \leq 0$ and this is a strong argument in favor of dS symmetry vs. Poincare and AdS ones. As argued in Sect. 1.3, quantum theory should start not from space-time but from a symmetry algebra. Therefore the choice of dS symmetry is natural and the cosmological constant problem does not exist. However, as noted in Secs. 1.4 and 1.5, the majority of physicists prefer to start from a flat space-time and treat Poincare symmetry as fundamental while dS one as emergent.

In contrast to the situation in Poincare and AdS invariant theories, the free mass operator in dS theory is not bounded below by the value of $m_1 + m_2$. The discussion in Sect. 3.6 shows that this property by no means implies that the theory is unphysical. Therefore if one has a choice between Poincare, AdS and dS symmetries then the only chance to describe gravity in a free theory is to choose dS symmetry.

3.2 IRs of the dS algebra

In view of the definition of elementary particle discussed in Sec. 2.4, we accept that, *by definition*, elementary particles in dS invariant theory are described by IRs of the dS algebra by Hermitian operators. For different reasons, there exists a vast literature not on such IRs but on UIRs of the dS group. References to this literature can be found e.g., in our papers [36, 35, 37] where we used the results on UIRs of the dS group for constructing IRs of the dS algebra by Hermitian operators. In this section we will describe the construction proceeding from an excellent description of UIRs of the dS group in a book by Mensky [46]. The final result is given by explicit expressions for the operators M^{ab} in Eqs. (3.16) and (3.17). The readers who are not interested in technical details can skip the derivation.

The elements of the SO(1,4) group will be described in the block form

$$g = \left\| \begin{array}{ccc} g_0^0 & \mathbf{a}^T & g_4^0 \\ \mathbf{b} & r & \mathbf{c} \\ g_0^4 & \mathbf{d}^T & g_4^4 \end{array} \right\| \tag{3.1}$$

where

$$\mathbf{a} = \left\| \begin{array}{c} a^1 \\ a^2 \\ a^3 \end{array} \right\|, \quad \mathbf{b}^T = \| \, b_1 \quad b_2 \quad b_3 \, \|, \quad r \in SO(3) \tag{3.2}$$

and the subscript ^{T} means a transposed vector.

UIRs of the SO(1,4) group belonging to the principle series of UIRs are induced from UIRs of the subgroup H (sometimes called "little group") defined as follows [46]. Each element of H can be uniquely represented as a product of elements of the subgroups SO(3), A and $\mathbf{T}$: $h = r\tau_A \mathbf{a_T}$ where

$$\tau_A = \left\| \begin{array}{ccc} cosh(\tau) & 0 & sinh(\tau) \\ 0 & 1 & 0 \\ sinh(\tau) & 0 & cosh(\tau) \end{array} \right\| \quad \mathbf{a_T} = \left\| \begin{array}{ccc} 1+\mathbf{a}^2/2 & -\mathbf{a}^T & \mathbf{a}^2/2 \\ -\mathbf{a} & 1 & -\mathbf{a} \\ -\mathbf{a}^2/2 & \mathbf{a}^T & 1-\mathbf{a}^2/2 \end{array} \right\| \tag{3.3}$$

The subgroup A is one-dimensional and the three-dimensional group $\mathbf{T}$ is the dS analog of the conventional translation group (see e.g., Ref. [46, 81]). We believe it should not cause misunderstandings when 1 is used in its usual meaning and when to denote the unit element of the SO(3) group. It should also be clear when r is a true element of the SO(3) group or belongs to the SO(3) subgroup of the SO(1,4) group. Note that standard UIRs of the Poincare group are induced from the little group, which is a semidirect product of SO(3) and four-dimensional translations and so the analogy between UIRs of the Poincare and dS groups is clear.

Let $r \rightarrow \Delta(r;\mathbf{s})$ be an UIR of the group SO(3) with the spin $\mathbf{s}$ and $\tau_A \rightarrow exp(im_{dS}\tau)$ be a one-dimensional UIR of the group A, where m_{dS} is a real parameter. Then UIRs of the group H used for inducing to the SO(1,4) group, have the form

$$\Delta(r\tau_A\mathbf{a_T};m_{dS},\mathbf{s}) = exp(im_{dS}\tau)\Delta(r;\mathbf{s}) \tag{3.4}$$

We will see below that m_{dS} has the meaning of the dS mass and therefore UIRs of the SO(1,4) group are defined by the mass and spin, by analogy with UIRs in Poincare invariant theory.

Let G=SO(1,4) and $X = G/H$ be the factor space (or coset space) of G over H. The notion of the factor space is known (see e.g., Refs. [82, 46]). Each element $x \in X$ is a class containing the elements $x_G h$ where $h \in H$, and $x_G \in G$ is a representative of the class x. The choice of representatives is not unique since if x_G is a representative of the class $x \in G/H$ then $x_G h_0$, where h_0 is an arbitrary element from H, also is a representative of the same class. It is known that X can be treated as a left G space. This means that if $x \in X$ then the action of the group G on X can be defined as follows: if $g \in G$ then gx is a class containing gx_G (it is easy to verify that such an action is correctly defined). Suppose that the choice of representatives is somehow fixed. Then $gx_G = (gx)_G(g,x)_H$ where $(g,x)_H$ is an element of H. This element is called a factor.

The explicit form of the operators M^{ab} depends on the choice of representatives in the space G/H. As explained in works on UIRs of the SO(1,4) group (see e.g., Ref. [46]), to obtain the possible closest analogy between UIRs of the SO(1,4) and Poincare groups, one should proceed as follows. Let $\mathbf{v}_L$ be a representative of the Lorentz group in the factor space SO(1,3)/SO(3) (strictly speaking, we should consider $SL(2,C)/SU(2)$). This space can be represented as the velocity hyperboloid with the Lorentz invariant measure

$$d\rho(\mathbf{v}) = d^3\mathbf{v}/v_0 \tag{3.5}$$

where $v_0 = (1+\mathbf{v}^2)^{1/2}$. Let $I \in SO(1,4)$ be a matrix which formally has the same form as the metric tensor η. One can show (see e.g., Refs. [46] for details) that $X = G/H$ can be represented as a union of three spaces, X_+, X_- and X_0 such that X_+ contains classes $\mathbf{v}_L h$, X_- contains classes $\mathbf{v}_L I h$ and X_0 has measure zero relative to the spaces X_+ and X_- (see also Sec. 3.4).

As a consequence, the space of UIR of the SO(1,4) group can be implemented as follows. If s is the spin of the particle under consideration, then we use $||...||$ to denote the norm in the space of UIR of the group SU(2) with the spin s. Then the space of UIR is the space of functions $\{f_1(\mathbf{v}), f_2(\mathbf{v})\}$ on two Lorentz hyperboloids with the range in the space of UIR of the group SU(2) with the spin s and such that

$$\int [||f_1(\mathbf{v})||^2 + ||f_2(\mathbf{v})||^2] d\rho(\mathbf{v}) < \infty \tag{3.6}$$

It is known that positive energy UIRs of the Poincare and AdS groups (associated with elementary particles) are implemented on an analog of X_+ while negative energy UIRs (associated with antiparticles) are implemented on an analog of X_-. Since the Poincare and AdS groups do not contain elements transforming these spaces to one another, the positive and negative energy UIRs are fully independent. At the same time, the dS group contains such elements (e.g. I [46, 81]) and for this reason its UIRs can be implemented only on the union of X_+ and X_-. Even this fact is a strong indication that UIRs of the dS group cannot be interpreted in the same way as UIRs of the Poincare and AdS groups.

A general construction of the operators M^{ab} follows. We first define right invariant measures on $G = SO(1,4)$ and H. It is known (see e.g. Ref. [82]) that for semisimple Lie groups (which is the case for the dS group), the right invariant measure is simultaneously the left invariant one. At the same time, the right invariant measure $d_R(h)$ on H is not the left invariant one, but has the property $d_R(h_0h) = \Delta(h_0)d_R(h)$, where the number function $h \to \Delta(h)$ on H is called the module of the group H. It is easy to show [46] that

$$\Delta(r\tau_A\mathbf{a_T}) = exp(-3\tau) \tag{3.7}$$

Let $d\rho(x)$ be a measure on $X = G/H$ compatible with the measures on G and H. This implies that the measure on G can be represented as $d\rho(x)d_R(h)$. Then one can show [46] that if X is a union of X_+ and X_- then the measure $d\rho(x)$ on each Lorentz hyperboloid coincides with that given by Eq. (3.5). Let the representation space be implemented as the space of functions $\varphi(x)$ on X with the range in the space of UIR of the SU(2) group such that

$$\int ||\varphi(x)||^2 d\rho(x) < \infty \tag{3.8}$$

Then the action of the representation operator $U(g)$ corresponding to $g \in G$ is

$$U(g)\varphi(x) = [\Delta((g^{-1},x)_H)]^{-1/2}\Delta((g^{-1},x)_H; m_{dS}, \mathbf{s})^{-1}\varphi(g^{-1}x) \tag{3.9}$$

One can directly verify that this expression defines a unitary representation. Its irreducibility can be proved in several ways (see e.g. Ref. [46]).

As noted above, if X is the union of X_+ and X_-, then the representation space can be implemented as in Eq. (3.4). Since we are interested in calculating only the explicit form of the operators M^{ab}, it suffices to consider only elements of $g \in G$ in an infinitely small vicinity of the unit element of the dS group. In that case one can calculate the action of representation operators on functions having the carrier in X_+ and X_- separately. Namely, as follows from Eq. (3.7), for such $g \in G$, one has to find the decompositions

$$g^{-1}\mathbf{v}_L = \mathbf{v}'_L r'(\tau')_A(\mathbf{a}')_\mathbf{T} \tag{3.10}$$

and

$$g^{-1}\mathbf{v}_L I = \mathbf{v}"_L I r"(\tau")_A(\mathbf{a}")_\mathbf{T} \tag{3.11}$$

where $r', r" \in SO(3)$. In this expressions it suffices to consider only elements of H belonging to an infinitely small vicinity of the unit element.

The problem of choosing representatives in the spaces SO(1,3)/SO(3) or SL(2.C)/SU(2) is well-known in standard theory. The most usual choice is such that $\mathbf{v}_L$ as an element of SL(2,C) is given by

$$\mathbf{v}_L = \frac{v_0 + 1 + \mathbf{v}\sigma}{\sqrt{2(1+v_0)}} \tag{3.12}$$

Then by using a known relation between elements of SL(2,C) and SO(1,3) we obtain that $\mathbf{v}_L \in SO(1,4)$ is represented by the matrix

$$\mathbf{v}_L = \left\| \begin{array}{ccc} v_0 & \mathbf{v}^T & 0 \\ \mathbf{v} & 1 + \mathbf{v}\mathbf{v}^T/(v_0+1) & 0 \\ 0 & 0 & 1 \end{array} \right\| \tag{3.13}$$

As follows from Eqs. (3.4) and (3.9), there is no need to know the expressions for $(\mathbf{a}')_{\mathbf{T}}$ and $(\mathbf{a}")_{\mathbf{T}}$ in Eqs. (3.10) and (3.11). We can use the fact [46] that if e is the five-dimensional vector with the components $(e^0 = 1, 0, 0, 0, e^4 = -1)$ and $h = r\tau_A \mathbf{a_T}$, then $he = exp(-\tau)e$ regardless of the elements $r \in SO(3)$ and $\mathbf{a_T}$. This makes it possible to easily calculate $(\mathbf{v}'_L, \mathbf{v}"_L, (\tau')_A, (\tau")_A)$ in Eqs. (3.10) and (3.11). Then one can calculate $(r', r")$ in these expressions by using the fact that the SO(3) parts of the matrices $(\mathbf{v}'_L)^{-1} g^{-1} \mathbf{v}_L$ and $(\mathbf{v}"_L)^{-1} g^{-1} \mathbf{v}_L$ are equal to r' and $r"$, respectively.

The relation between the operators $U(g)$ and M^{ab} follows. Let L_{ab} be the basis elements of the Lie algebra of the dS group. These are the matrices with the elements

$$(L_{ab})^c_d = \delta^c_d \eta_{bd} - \delta^c_b \eta_{ad} \tag{3.14}$$

They satisfy the commutation relations

$$[L_{ab}, L_{cd}] = \eta_{ac} L_{bd} - \eta_{bc} L_{ad} - \eta_{ad} L_{bc} + \eta_{bd} L_{ac} \tag{3.15}$$

Comparing Eqs. (1.4) and (3.15) it is easy to conclude that the M^{ab} should be the representation operators of $-iL^{ab}$. Therefore if $g = 1 + \omega_{ab} L^{ab}$, where a sum over repeated indices is assumed and the ω_{ab} are such infinitely small parameters that $\omega_{ab} = -\omega_{ba}$ then $U(g) = 1 + i\omega_{ab} M^{ab}$.

We are now in position to write down the final expressions for the operators M^{ab}. Their action on functions with the carrier in X_+ has the form

$$\begin{aligned} &\mathbf{M}^{(+)} = l(\mathbf{v}) + \mathbf{s}, \quad \mathbf{N}^{(+)} = -iv_0 \frac{\partial}{\partial \mathbf{v}} + \frac{\mathbf{s} \times \mathbf{v}}{v_0 + 1} \\ &\mathbf{B}^{(+)} = m_{dS}\mathbf{v} + i[\frac{\partial}{\partial \mathbf{v}} + \mathbf{v}(\mathbf{v}\frac{\partial}{\partial \mathbf{v}}) + \frac{3}{2}\mathbf{v}] + \frac{\mathbf{s} \times \mathbf{v}}{v_0 + 1} \\ &\mathcal{E}^{(+)} = m_{dS} v_0 + iv_0(\mathbf{v}\frac{\partial}{\partial \mathbf{v}} + \frac{3}{2}) \end{aligned} \tag{3.16}$$

where $\mathbf{J} = \{M^{23}, M^{31}, M^{12}\}$, $\mathbf{N} = \{M^{01}, M^{02}, M^{03}\}$, $\mathbf{B} = \{M^{41}, M^{42}, M^{43}\}$, $\mathbf{s}$ is the spin operator, $\mathbf{l}(\mathbf{v}) = -i\mathbf{v} \times \partial/\partial\mathbf{v}$ and $\mathcal{E} = M^{40}$. At the same time, the action on functions with the carrier in X_- is given by

$$\mathbf{J}^{(-)} = l(\mathbf{v}) + \mathbf{s}, \quad \mathbf{N}^{(-)} = -iv_0\frac{\partial}{\partial\mathbf{v}} + \frac{\mathbf{s}\times\mathbf{v}}{v_0+1}$$
$$\mathbf{B}^{(-)} = -m_{dS}\mathbf{v} - i[\frac{\partial}{\partial\mathbf{v}} + \mathbf{v}(\mathbf{v}\frac{\partial}{\partial\mathbf{v}}) + \frac{3}{2}\mathbf{v}] - \frac{\mathbf{s}\times\mathbf{v}}{v_0+1}$$
$$\mathcal{E}^{(-)} = -m_{dS}v_0 - iv_0(\mathbf{v}\frac{\partial}{\partial\mathbf{v}} + \frac{3}{2}) \tag{3.17}$$

Note that the expressions for the action of the Lorentz algebra operators on X_+ and X_- are the same and they coincide with the corresponding expressions for IRs of the Poincare algebra. At the same time, the expressions for the action of the operators $M^{4\mu}$ on X_+ and X_- differ by sign.

In deriving Eqs. (3.16) and (3.17) we have used only the commutation relations (1.4), no approximations have been made and the results are exact. In particular, the dS space, the cosmological constant and the Riemannian geometry have not been involved at all. Nevertheless, the expressions for the representation operators is all we need to have the maximum possible information in quantum theory. As shown in the literature (see e.g. Ref. [46]), the above construction of IRs applies to IRs of the principle series where m_{dS} is a nonzero real parameter. Therefore such IRs are called massive.

A problem arises how m_{dS} is related to the standard particle mass m in Poincare invariant theory. In view of the contraction procedure described in Sec. 1.3, one can assume that $m_{dS} > 0$ and define $m = m_{dS}/R$, $\mathbf{P} = \mathbf{B}/R$ and $E = \mathcal{E}/R$. The set of operators $(E, \mathbf{P})$ is the Lorentz vector since its components can be written as $M^{4\nu}/R$ ($\nu = 0, 1, 2, 3$). Then, as follows from Eqs. (1.4), in the limit when $R \to \infty$, $m_{dS} \to \infty$ but m_{ds}/R is finite, one obtains from Eq. (3.16) a standard positive energy representation of the Poincare algebra for a particle with the mass m such that $\mathbf{P} = m\mathbf{v}$ is the particle momentum and $E = mv_0$ is the particle energy. Analogously one obtains a negative energy representation from Eq. (3.17). Therefore m is the standard mass in Poincare invariant theory and the operators of the Lorentz algebra $(\mathbf{N}, \mathbf{J})$ have the same form for the Poincare and dS algebras.

In Sect. 1.4 we have argued that fundamental physical theory should not contain dimensionful parameters at all. In this connection it is interesting to note that the de Sitter mass m_{dS} is a ratio of the radius of the World R to the Compton wave length of the particle under consideration. Therefore even for elementary particles the de Sitter masses are very large. For example, if R is of the order of $10^{26}m$ then the de Sitter masses of the electron, the Earth and the Sun are of the order of 10^{39}, 10^{93} and 10^{99}, respectively. The fact that even the dS mass of the electron is so large might be an indication that the electron is not a true elementary particle. Moreover, the present upper level for the photon mass is $10^{-18}ev$ which seems to be an extremely tiny quantity. However, the corresponding dS mass is of the order of 10^{15} and so even the mass which is treated as extremely small in Poincare invariant theory might be very large in dS invariant theory.

The operator $\mathbf{N}$ contains $i\partial/\partial\mathbf{v}$ which is proportional to the standard coordinate operator $i\partial/\partial\mathbf{p}$. The factor v_0 in $\mathbf{N}$ is needed for Hermiticity since the volume element is given by Eq. (3.5). Such a construction can be treated as a relativistic generalization of standard coordinate operator and then the orbital part of $\mathbf{N}$ is proportional to the Newton-Wigner position operator [20]. However, as shown in Chap. 2, this operator does not satisfy all the requirements for the coordinate operator.

In Poincare invariant theory the operator $I_{2P} = E^2 - \mathbf{P}^2$ is the Casimir operator, *i.e.*, it commutes with all the representation operators. According to the known Schur lemma in representation theory, all elements in the space of IR are eigenvectors of the Casimir operators with the same eigenvalue. In particular, they are the eigenvectors of the operator I_{2P} with the eigenvalue m^2. As follows from Eq. (1.4), in the dS case the Casimir operator of the second order is

$$I_2 = -\frac{1}{2}\sum_{ab} M_{ab}M^{ab} = \mathcal{E}^2 + \mathbf{N}^2 - \mathbf{B}^2 - \mathbf{J}^2 \tag{3.18}$$

and a direct calculation shows that for the operators (3.16) and (3.17) the numerical value of I_2 is $m_{dS}^2 - s(s+1) + 9/4$. In Poincare invariant theory the value of the spin is related to the Casimir operator of the fourth order which can be constructed from the Pauli-Lubanski vector. An analogous construction exists in dS invariant theory but we will not dwell on this.

3.3 Absence of Weyl particles in dS invariant theory

According to Standard Model, only massless Weyl particles can be fundamental elementary particles in Poincare invariant theory. Therefore a problem arises whether there exist analogs of Weyl particles in dS invariant theory. In Poincare invariant theory, Weyl particles are characterized not only by the condition that their mass is zero but also by the condition that they have a definite helicity. Several authors investigated dS and AdS analogs of Weyl particles proceeding from covariant equations on the dS and AdS spaces, respectively. For example, the authors of Ref. [83] show that Weyl particles arise only when dS or AdS symmetries are broken to Lorentz symmetry. At the level of IRs, the existence of analogs of Weyl particles is known in the AdS case. In Ref. [41] we investigated such analogs by using the results of Refs. [80] for standard IRs of the AdS algebra (i.e. IRs over the field of complex numbers) and the results of Ref. [84] for IRs of the AdS algebra over a Galois field (see also Sec. 8.3 of the present work). In the standard case the minimum value of the AdS energy for massless IRs with positive energy is $E_{min} = 1 + s$. In contrast to the situation in Poincare invariant theory, where massless particles cannot be in the rest state, massless particles in the AdS theory do have rest states and the value of the z projection of the spin in such states can be $-s, -s+1, ..., s$ as usual. However, for any value of the energy greater than E_{min}, the spin state is characterized only by helicity, which can take the values either s or $-s$, *i.e.*, we have the same result as in Poincare invariant theory. In contrast to IRs of the Poincare and dS algebra, IRs describing particles in AdS theory belong to the discrete series of IRs and the energy spectrum is discrete: $E = E_{min}, E_{min} + 1, ..., \infty$. Therefore, strictly speaking, rest states do not have measure zero. Nevertheless, the probability that the energy is exactly E_{min} is extremely small and therefore there exists a correspondence between Weyl particles in Poincare and AdS theories.

In Poincare invariant theory, IRs describing Weyl particles can be constructing by analogy with massive IRs but the little group is now E(2) instead of SO(3) (see e.g. Sec. 2.5 in the textbook [2]). The matter is that the representation operators of the SO(3) group transform rest states into themselves but for massless particles there are no rest states. However, there exists another way of getting massless IRs: one can choose the variables for massive IRs in such a way that the operators of massless IRs can be directly obtained from the operators of massive IRs in the limit $m \to 0$. This construction has been described by several authors (see e.g. Refs. [85, 86, 56] and references therein) and the main stages follow. First, instead of the $(0, 1, 2, 3)$ components of vectors, we work with the so called light front components $(+, -, 1, 2)$ where $v^{\pm} = (v^0 \pm v^3)/\sqrt{2}$ and analogously for other vectors. We choose $(v^+, \mathbf{v}_\perp)$ as three independent components of the 4-velocity vector, where $\mathbf{v}_\perp = (v_x, v_y)$. In these variables the measure (3.5) on the Lorentz hyperboloid becomes $d\rho(v^+, \mathbf{v}_\perp) = dv^+ d\mathbf{v}_\perp / v^+$. Instead of Eq. (3.12) we now choose representatives of the SL(2,C)/SU(2) classes as

$$v_L = \frac{1}{(v_0 + v_z)^{1/2}} \left\| \begin{array}{cc} v_0 + v_z & 0 \\ v_x + iv_y & 1 \end{array} \right\| \qquad (3.19)$$

and by using the relation between the groups SL(2,C) and SO(1,3) we obtain that the form of this representative in the Lorentz group is

$$v_L = \left\| \begin{array}{cccc} \sqrt{2}v^+ & 0 & 0 & 0 \\ \frac{\mathbf{v}_\perp^2}{\sqrt{2}v^+} & \frac{1}{\sqrt{2}v^+} & \frac{v_x}{v^+} & \frac{v_y}{v^+} \\ \sqrt{2}v_x & 0 & 1 & 0 \\ \sqrt{2}v_y & 0 & 0 & 1 \end{array} \right\| \qquad (3.20)$$

where the raws and columns are in the order $(+, -, x, y)$.

By using the scheme described in the preceding section, we can now calculate the explicit form of the representation operators of the Lorentz algebra. In this scheme the form of these operators in the IRs of the Poincare and dS algebras is the same and in the case of the dS algebra the action is the same for states with the carrier in X_+ and X_-. The results of calculations are:

$$M^{+-} = iv^+\frac{\partial}{\partial v^+}, \quad M^{+j} = iv^+\frac{\partial}{\partial v^j}, \quad M^{12} = l_z(\mathbf{v}_\perp) + s_z$$
$$M^{-j} = -i(v^j\frac{\partial}{\partial v^+} + v^-\frac{\partial}{\partial v^j}) - \frac{\epsilon_{jl}}{v^+}(s^l + v^l s_z) \tag{3.21}$$

where a sum over $j, l = 1, 2$ is assumed and ϵ_{jl} has the components $\epsilon_{12} = -\epsilon_{21} = 1$, $\epsilon_{11} = \epsilon_{22} = 0$. In Poincare invariant theories one can define the standard four-momentum $p = mv$ and choose $(p^+, \mathbf{p}_\perp)$ as independent variables. Then the expressions in Eq. (3.21) can be rewritten as

$$M^{+-} = ip^+\frac{\partial}{\partial p^+}, \quad M^{+j} = ip^+\frac{\partial}{\partial p^j}, \quad M^{12} = l_z(\mathbf{p}_\perp) + s_z$$
$$M^{-j} = -i(p^j\frac{\partial}{\partial p^+} + p^-\frac{\partial}{\partial p^j}) - \frac{\epsilon_{jl}}{p^+}(ms^l + p^l s_z) \tag{3.22}$$

In dS invariant theory we can work with the same variables if m is defined as m_{dS}/R.

As seen from Eqs. (3.22), only the operators M^{-j} contain a dependence on the operators s_x and s_y but this dependence disappears in the limit $m \to 0$. In this limit the operator s_z can be replaced by its eigenvalue λ which now has the meaning of helicity. In Poincare invariant theory the four-momentum operators P^μ are simply the operators of multiplication by p^μ and therefore massless particles are characterized only by one constant—helicity.

In dS invariant theory one can calculate the action of the operators $M^{4\mu}$ by analogy with the calculation in the preceding section. The actions of these operators on states with the carrier in X_+ and X_- differ only by sign and the result for the actions on states with the carrier in X_+ is

$$M^{4-} = m_{dS}v^- + i[v^-(v^+\frac{\partial}{\partial v^+} + v^j\frac{\partial}{\partial v^j} + \frac{3}{2}) - \frac{\partial}{\partial v^+}] + \frac{1}{v^+}\epsilon_{jl}v^j s^l$$
$$M^{4j} = m_{dS}v^j + i[v^j(v^+\frac{\partial}{\partial v^+} + v^l\frac{\partial}{\partial v^l} + \frac{3}{2}) + \frac{\partial}{\partial v^j}] - \epsilon_{jl}s^l$$
$$M^{4+} = m_{dS}v^+ + iv^+(v^+\frac{\partial}{\partial v^+} + v^j\frac{\partial}{\partial v^j} + \frac{3}{2}) \tag{3.23}$$

If we define $m = m_{dS}/R$ and $p^\mu = mv^\mu$ then for the operators P^μ we have

$$P^- = p^- + \frac{ip^-}{mR}(p^+\frac{\partial}{\partial p^+} + p^j\frac{\partial}{\partial p^j} + \frac{3}{2}) - \frac{im}{R}\frac{\partial}{\partial p^+} + \frac{1}{Rp^+}\epsilon_{jl}p^j s^l$$
$$P^j = p^j + \frac{ip^j}{mR}(p^+\frac{\partial}{\partial p^+} + p^l\frac{\partial}{\partial p^l} + \frac{3}{2}) + \frac{im}{R}\frac{\partial}{\partial p^j} - \frac{1}{R}\epsilon_{jl}s^l$$
$$P^+ = p^+ + \frac{ip^+}{mR}(p^+\frac{\partial}{\partial p^+} + p^j\frac{\partial}{\partial p^j} + \frac{3}{2}) \tag{3.24}$$

Then it is clear that in the formal limit $R \to \infty$ we obtain the standard Poincare result. However, when R is finite, the dependence of the operators P^μ on s_x and s_y does not disappear. Moreover, in this case we cannot take the limit $m \to 0$. Therefore we conclude that in dS theory there are no Weyl particles, at least in the case when elementary particles are described by IRs of the principle series. Mensky conjectured [46] that massless particles in dS invariant theory might correspond to IRs of the discrete series with $-im_{dS} = 1/2$ but this possibility has not been investigated. In any case, in contrast to the situation in Poincare invariant theory, the limit of massive IRs when $m \to 0$ does not give Weyl particles and moreover, this limit does not exist.

3.4 Other implementations of IRs

In this section we briefly describe two more implementations of IRs of the dS algebra. The first one is based on the fact that since SO(1,4)=SO(4)$A\mathbf{T}$ and H=SO(3)$A\mathbf{T}$ [46], there also exists a choice of representatives which is probably even more natural than those described above. Namely, we can choose as representatives the elements from the coset space SO(4)/SO(3). Since the universal covering group for SO(4) is SU(2)×SU(2) and for SO(3) — SU(2), we can choose as representatives the elements of the first multiplier in the product SU(2)×SU(2). Elements of SU(2) can be represented by the points $u = (\mathbf{u}, u_4)$ of the three-dimensional sphere S^3 in the four-dimensional space as $u_4 + i\sigma\mathbf{u}$ where σ are the Pauli matrices and $u_4 = \pm(1-\mathbf{u}^2)^{1/2}$ for the upper and lower hemispheres, respectively. Then the calculation of the operators is similar to that described above and the results follow. The Hilbert space is now the space of functions $\varphi(u)$ on S^3 with the range in the space of the IR of the su(2) algebra with the spin s and such that

$$\int ||\varphi(u)||^2 du < \infty \tag{3.25}$$

where du is the SO(4) invariant volume element on S^3. The explicit calculation shows that in this case the operators have the form

$$\mathbf{J} = l(\mathbf{u}) + \mathbf{s}, \quad \mathbf{B} = iu_4\frac{\partial}{\partial\mathbf{u}} - \mathbf{s}, \quad \mathcal{E} = (m_{dS} + 3i/2)u_4 + iu_4\mathbf{u}\frac{\partial}{\partial\mathbf{u}}$$
$$\mathbf{N} = -i[\frac{\partial}{\partial\mathbf{u}} - \mathbf{u}(\mathbf{u}\frac{\partial}{\partial\mathbf{u}})] + (m_{dS} + 3i/2)\mathbf{u} - \mathbf{u}\times\mathbf{s} + u_4\mathbf{s} \tag{3.26}$$

Since Eqs. (3.6), (3.16) and (3.17) on one hand and Eqs. (3.25) and (3.26) on the other are the different implementations of one and the same representation, there exists a unitary operator transforming functions $f(v)$ into $\varphi(u)$ and operators (3.16,3.17) into operators (3.26). For example in the spinless case the operators (3.16) and (3.26) are related to each other by a unitary transformation

$$\varphi(u) = exp(-im_{dS}lnv_0)v_0^{3/2}f(v) \tag{3.27}$$

where the relation between the points of the upper hemisphere and X_+ is $\mathbf{u} = \mathbf{v}/v_0$ and $u_4 = (1 - \mathbf{u}^2)^{1/2}$. The relation between the points of the lower hemisphere and X_- is $\mathbf{u} = -\mathbf{v}/v_0$ and $u_4 = -(1 - \mathbf{u}^2)^{1/2}$.

The equator of S^3 where $u_4 = 0$ corresponds to X_0 and has measure zero with respect to the upper and lower hemispheres. For this reason one might think that it is of no interest for describing particles in dS theory. Nevertheless, an interesting observation is that while none of the components of u has the magnitude greater than unity, the set X_0 in terms of velocities is characterized by the condition that $|\mathbf{v}|$ is infinitely large and therefore the standard Poincare momentum $\mathbf{p} = m\mathbf{v}$ is infinitely large too. This poses a question whether $\mathbf{p}$ always has a physical meaning. From mathematical point of view Eq. (3.26) might seem more convenient than Eqs. (3.16) and (3.17) since S^3 is compact and there is no need to break it into the upper and lower hemispheres. In addition, Eq. (3.26) is an explicit implementation of the idea that since in dS invariant theory all the variables (x^1, x^2, x^3, x^4) are on equal footing and so(4) is the maximal compact kinematical algebra, the operators $\mathbf{M}$ and $\mathbf{B}$ do not depend on m_{dS}. However, those expressions are not convenient for investigating Poincare approximation since the Lorentz boost operators $\mathbf{N}$ depend on m_{dS}.

Finally, we describe an implementation of IRs based on the explicit construction of the basis in the representation space. This construction is based on the method of su(2)×su(2) shift operators, developed by Hughes [87] for constructing UIRs of the group SO(5). It will be convenient for us to deal with the set of operators $(\mathbf{J}', \mathbf{J}'', R_{ij})$ $(i, j = 1, 2)$ instead of M^{ab}. Here $\mathbf{J}'$ and $\mathbf{J}$" are two independent su(2) algebras (*i.e.*, $[\mathbf{J}', \mathbf{J}''] = 0$). In each of them one chooses as the basis the operators (J_+, J_-, J_3) such that $J_1 = J_+ + J_-$, $J_2 = -\imath(J_+ - J_-)$ and the commutation relations have the form

$$[J_3, J_+] = 2J_+, \quad [J_3, J_-] = -2J_-, \quad [J_+, J_-] = J_3 \tag{3.28}$$

The commutation relations of the operators $\mathbf{J}'$ and $\mathbf{J}$" with R_{ij} have the form

$$\begin{aligned}
&[J_3', R_{1j}] = R_{1j}, \quad [J_3', R_{2j}] = -R_{2j}, \quad [J_3'', R_{i1}] = R_{i1}, \\
&[J_3'', R_{i2}] = -R_{i2}, \quad [J_+', R_{2j}] = R_{1j}, \quad [J_+'', R_{i2}] = R_{i1}, \\
&[J_-', R_{1j}] = R_{2j}, \quad [J_-'', R_{i1}] = R_{i2}, \quad [J_+', R_{1j}] = \\
&[J_+'', R_{i1}] = [J_-', R_{2j}] = [J_-'', R_{i2}] = 0
\end{aligned} \tag{3.29}$$

and the commutation relations of the operators R_{ij} with each other have the form

$$\begin{aligned}
&[R_{11}, R_{12}] = 2J_+', \quad [R_{11}, R_{21}] = 2J_+'', \\
&[R_{11}, R_{22}] = -(J_3' + J_3''), \quad [R_{12}, R_{21}] = J_3' - J_3'' \\
&[R_{11}, R_{22}] = -2J_-'', \quad [R_{21}, R_{22}] = -2J_-'
\end{aligned} \tag{3.30}$$

The relation between the sets $(\mathbf{J}', \mathbf{J}", R_{ij})$ and M^{ab} is given by

$$\begin{aligned}
&\mathbf{J} = (\mathbf{J}' + \mathbf{J}'')/2, \quad \mathbf{B} = (\mathbf{J}' - \mathbf{J}'')/2, \quad M_{01} = i(R_{11} - R_{22})/2,\\
&M_{02} = (R_{11} + R_{22})/2, \quad M_{03} = -i(R_{12} + R_{21})/2,\\
&M_{04} = (R_{12} - R_{21})/2
\end{aligned} \tag{3.31}$$

Then it is easy to see that Eq. (1.4) follows from Eqs. (3.29–3.31) and *vice versa*.

Consider the space of maximal $su(2) \times su(2)$ vectors, *i.e.*, such vectors x that $J'_+x = J''_+x = 0$. Then from Eqs. (3.29) and (3.30) it follows that the operators

$$\begin{aligned}
&A^{++} = R_{11}, \quad A^{+-} = R_{12}(J''_3 + 1) - J''_- R_{11}, \quad A^{-+} = R_{21}(J'_3 + 1) - J'_- R_{11},\\
&A^{--} = -R_{22}(J'_3 + 1)(J''_3 + 1) + J''_- R_{21}(J'_3 + 1) +\\
&J'_- R_{12}(J''_3 + 1) - J'_- J''_- R_{11}
\end{aligned} \tag{3.32}$$

act invariantly on this space. The notations are related to the property that if x^{kl} $(k, l > 0)$ is the maximal su(2)×su(2) vector and simultaneously the eigenvector of operators J'_3 and J_3" with the eigenvalues k and l, respectively, then $A^{++}x^{kl}$ is the eigenvector of the same operators with the values $k+1$ and $l+1$, $A^{+-}x^{kl}$ - the eigenvector with the values $k+1$ and $l-1$, $A^{-+}x^{kl}$ - the eigenvector with the values $k-1$ and $l+1$ and $A^{--}x^{kl}$ - the eigenvector with the values $k-1$ and $l-1$.

The basis in the representation space can be explicitly constructed assuming that there exists a vector e^0 which is the maximal su(2)×su(2) vector such that

$$J'_3 e_0 = 0, \quad J''_3 e_0 = s e_0, \quad A^{--} e_0 = A^{-+} e_0 = 0, \quad I_2 e^0 = [m_{dS}^2 - s(s+1) + 9/4] e^0 \tag{3.33}$$

Then, as shown in Ref. [35], the full basis of the representation space consists of vectors

$$e_{ij}^{nr} = (J'_-)^i (J''_-)^j (A^{++})^n (A^{+-})^r e^0 \tag{3.34}$$

where $n = 0, 1, 2, ..., r$ can take only the values $0, 1, ..., 2s$ and for the given n and s, i can take the values $0, 1, ..., n+r$ and j can take the values $0, 1, ..., n+2s-r$.

These results show that IRs of the dS algebra can be constructed purely algebraically without involving analytical methods of the theory of UIRs of the dS group. As shown in Ref. [35], this implementation is convenient for generalizing standard quantum theory to a quantum theory over a Galois field. In Chap. 4 we consider in detail the algebraic construction of IRs in the spinless case and the results are applied to gravity.

3.5 Physical interpretation of IRs of the dS algebra

In Secs. 3.2–3.4 we discussed mathematical properties of IRs of the dS algebra. In particular it has been noted that they are implemented on two Lorentz hyperboloids, not one as IRs of the Poincare algebra. Therefore the number of states in IRs of the dS algebra is twice as big as in IRs of the Poincare algebra. A problem arises whether this is compatible with a requirement that any dS invariant theory should become a Poincare invariant one in the formal limit $R \to \infty$. Although there exists a wide literature on IRs of the dS group and algebra, their physical interpretation has not been widely discussed. Probably one of the reasons is that physicists working on dS QFT treat fields as more fundamental objects than particles (although the latter are observables while the former are not).

In his book [46] Mensky notes that, in contrast to IRs of the Poincare and AdS groups, IRs of the dS group characterized by m_{dS} and $-m_{dS}$ are unitarily equivalent and therefore the energy sign cannot be used for distinguishing particles and antiparticles. He proposes an interpretation where a particle and its antiparticle are described by the same IRs but have different space-time descriptions (defined by operators intertwining IRs with representations induced from the Lorentz group). Mensky shows that in the general case his two solutions still cannot be interpreted as a particle and its antiparticle, respectively, since they are nontrivial linear combinations of functions with different energy signs. However, such an interpretation is recovered in Poincare approximation.

In view of the above discussion, it is desirable to give an interpretation of IRs which does not involve space-time. In Ref. [37] we have proposed an interpretation such that one IR describes a particle and its antiparticle simultaneously. In this section this analysis is extended.

3.5.1 Problems with physical interpretation of IRs

Consider first the case when the quantity m_{dS} is very large. Then, as follows from Eqs. (3.16) and (3.17), the action of the operators $M^{4\mu}$ on states localized on X_+ or X_- can be approximately written as $\pm m_{dS} v^\mu$, respectively. Therefore a question arises whether the standard Poincare energy E can be defined as $E = M_{04}/R$. Indeed, with such a definition, states localized on X_+ will have a positive energy while states localized on X_- will have a negative energy. Then a question arises whether this is compatible with the standard interpretation of IRs, according to which the following requirements should be satisfied:

Standard-Interpretation Requirements: Each element of the full representation space represents a possible physical state for the given elementary particle. The representation describing a system of N free elementary particles is the tensor product of the corresponding single-particle representations.

Recall that the operators of the tensor product are given by sums of the corresponding single-particle operators. For example, if $\mathcal{E}^{(1)}$ is the operator $\mathcal{E}$ for particle 1 and $\mathcal{E}^{(2)}$ is the operator $\mathcal{E}$ for particle 2 then the operator $\mathcal{E}$ for the free system $\{12\}$ is given by $\mathcal{E}^{(12)} = \mathcal{E}^{(1)} + \mathcal{E}^{(2)}$. Here it is assumed that the action of the operator $\mathcal{E}^{(j)}$ $(j = 1, 2)$ in the two-particle space is defined as follows. It acts according to Eq. (3.16) or (3.17) over its respective variables while over the variables of the other particle it acts as the identity operator.

One could try to satisfy the standard interpretation as follows.

A) Assume that in Poincare approximation the standard energy should be defined as $E = \pm\mathcal{E}/R$ where the plus sign should be taken for the states with the carrier in X_+ and as the minus sign—for the states with the carrier in X_-. Then the energy will always be positive definite.

B) One might say that the choice of the energy sign is only a matter of convention. Indeed, to measure the energy of a particle with the mass m one has to measure its momentum $\mathbf{p}$ and then the energy can be defined not only as $(m^2 + \mathbf{p}^2)^{1/2}$ but also as $-(m^2 + \mathbf{p}^2)^{1/2}$. In that case the standard energy in the Poincare approximation could be defined as $E = \mathcal{E}/R$ regardless of whether the carrier of the given state is in X_+ or X_-.

It is easy to see that either of the above possibilities is incompatible with Standard-Interpretation Requirements. Consider, for example, a system of two free particles in the case when m_{dS} is very large. Then with a high accuracy the operators $\mathcal{E}/R$ and $\mathbf{B}/R$ can be chosen diagonal simultaneously.

Let us first assume that the energy should be treated according to B). Then a system of two free particles with the equal masses can have the same quantum numbers as the vacuum (for example, if the first particle has the energy $E_0 = (m^2 + \mathbf{p}^2)^{1/2}$ and momentum $\mathbf{p}$ while the second one has the energy $-E_0$ and the momentum $-\mathbf{p}$) what obviously contradicts experiment. For this and other reasons it is known that in Poincare invariant theory the particles should have the same energy sign. Analogously, if the single-particle energy is treated according to A) then the result for the two-body energy of a particle-antiparticle system will contradict experiment.

We conclude that IRs of the dS algebra cannot be interpreted in the standard way since such an interpretation is physically meaningless even in Poincare approximation. The above discussion indicates that the problem we have is similar to that with the interpretation of the fact that the Dirac equation has solutions with both, positive and negative energies.

As already noted, in Poincare and AdS theories there exist positive energy IRs implemented on the upper hyperboloid and negative energy IRs implemented on the lower hyperboloid. In the latter case Standard-Interpretation Requirements are not satisfied for the reasons discussed above. However, we cannot declare such IRs unphysical and throw them away. In QFT quantum fields necessarily contain both types of IRs such that positive energy IRs are associated with particles while negative energy IRs are associated with antiparticles. Then the energy of antiparticles can be made positive after proper second quantization. In view of this observation, we will investigate whether IRs of the dS algebra can be interpreted in such a way that one IR describes a particle and its antiparticle simultaneously such that states localized on X_+ are associated with a particle while states localized on X_- are associated with its antiparticle.

By using Eq. (3.6), one can directly verify that the operators (3.16) and (3.17) are Hermitian if the scalar product in the space of IR is defined as follows. Since the functions $f_1(\mathbf{v})$ and $f_2(\mathbf{v})$ in Eq. (3.6) have the range in the space of IR of the su(2) algebra with the spin s, we can replace them by the sets of functions $f_1(\mathbf{v}, j)$ and $f_2(\mathbf{v}, j)$, respectively, where $j = -s, -s+1, ..., s$. Moreover, we can combine these functions into one function $f(\mathbf{v}, j, \epsilon)$ where the variable ϵ can take only two values, say +1 or -1, for the components having the carrier in X_+ or X_-, respectively. If now $\varphi(\mathbf{v}, j, \epsilon)$ and $\psi(\mathbf{v}, j, \epsilon)$ are two elements of our Hilbert space, their scalar product is defined as

$$(\varphi, \psi) = \sum_{j,\epsilon} \int \varphi(\mathbf{v}, j, \epsilon)^* \psi(\mathbf{v}, j, \epsilon) d\rho(\mathbf{v}) \tag{3.35}$$

where the subscript * applied to scalar functions means the usual complex conjugation.

At the same time, we use * to denote the operator adjoint to a given one. Namely, if A is the operator in our Hilbert space then A^* means the operator such that

$$(\varphi, A\psi) = (A^*\varphi, \psi) \tag{3.36}$$

for all such elements φ and ψ that the left hand side of this expression is defined.

Even in the case of the operators (3.16) and (3.17) we can formally treat them as integral operators with some kernels. Namely, if $A\varphi = \psi$, we can treat this relation as

$$\sum_{j',\epsilon'} \int A(\mathbf{v}, j, \epsilon; \mathbf{v}', j', \epsilon')\varphi(\mathbf{v}', j', \epsilon') d\rho(\mathbf{v}') = \psi(\mathbf{v}, j, \epsilon) \tag{3.37}$$

where in the general case the kernel $A(\mathbf{v}, j, \epsilon; \mathbf{v}', j', \epsilon')$ of the operator A is a distribution.

As follows from Eqs. (3.35–3.37), if $B = A^*$ then the relation between the kernels of these operators is

$$B(\mathbf{v}, j, \epsilon; \mathbf{v}', j', \epsilon') = A(\mathbf{v}', j', \epsilon'; \mathbf{v}, j, \epsilon)^* \tag{3.38}$$

In particular, if the operator A is Hermitian then

$$A(\mathbf{v}, j, \epsilon; \mathbf{v}', j', \epsilon')^* = A(\mathbf{v}', j', \epsilon'; \mathbf{v}, j, \epsilon) \tag{3.39}$$

and if, in addition, its kernel is real then the kernel is symmetric, *i.e.*,

$$A(\mathbf{v}, j, \epsilon; \mathbf{v}', j', \epsilon') = A(\mathbf{v}', j', \epsilon'; \mathbf{v}, j, \epsilon) \tag{3.40}$$

In particular, this property is satisfied for the operators $m_{dS}v_0$ and $m_{dS}\mathbf{v}$ in Eqs. (3.16) and (3.17). At the same time, the operators

$$l(\mathbf{v}) \quad -iv_0\frac{\partial}{\partial \mathbf{v}} \quad -i[\frac{\partial}{\partial \mathbf{v}} + \mathbf{v}(\mathbf{v}\frac{\partial}{\partial \mathbf{v}}) + \frac{3}{2}\mathbf{v}] \quad -iv_0(\mathbf{v}\frac{\partial}{\partial \mathbf{v}} + \frac{3}{2}) \tag{3.41}$$

which are present in Eqs. (3.16) and (3.17), are Hermitian but have imaginary kernels. Therefore, as follows from Eq. (3.39), their kernels are antisymmetric:

$$A(\mathbf{v},j,\epsilon;\mathbf{v}',j',\epsilon') = -A(\mathbf{v}',j',\epsilon';\mathbf{v},j,\epsilon) \tag{3.42}$$

In standard approach to quantum theory, the operators of physical quantities act in the Fock space of the given system. Suppose that the system consists of free particles and their antiparticles. Strictly speaking, in our approach it is not clear yet what should be treated as a particle or antiparticle. The considered IRs of the dS algebra describe objects such that $(\mathbf{v},j,\epsilon)$ is the full set of their quantum numbers. Therefore we can define the annihilation and creation operators $(a(\mathbf{v},j,\epsilon),a(\mathbf{v},j,\epsilon)^*)$ for these objects. If the operators satisfy the anticommutation relations then we require that

$$\{a(\mathbf{v},j,\epsilon),a(\mathbf{v}',j',\epsilon')^*\} = \delta_{jj'}\delta_{\epsilon\epsilon'}v_0\delta^{(3)}(\mathbf{v}-\mathbf{v}') \tag{3.43}$$

while in the case of commutation relations

$$[a(\mathbf{v},j,\epsilon),a(\mathbf{v}',j',\epsilon')^*] = \delta_{jj'}\delta_{\epsilon\epsilon'}v_0\delta^{(3)}(\mathbf{v}-\mathbf{v}') \tag{3.44}$$

In the first case, any two a-operators or any two a^* operators anticommute with each other while in the second case they commute with each other.

The problem of second quantization can now be formulated such that IRs should be implemented as Fock spaces, i.e. states and operators should be expressed in terms of the (a,a^*) operators. A possible implementation follows. We define the vacuum state Φ_0 such that it has a unit norm and satisfies the requirement

$$a(\mathbf{v},j,\epsilon)\Phi_0 = 0 \quad \forall\ \mathbf{v},j,\epsilon \tag{3.45}$$

The image of the state $\varphi(\mathbf{v},j,\epsilon)$ in the Fock space is defined as

$$\varphi_F = \sum_{j,\epsilon}\int \varphi(\mathbf{v},j,\epsilon)a(\mathbf{v},j,\epsilon)^*d\rho(\mathbf{v})\Phi_0 \tag{3.46}$$

and the image of the operator with the kernel $A(\mathbf{v},j,\epsilon;\mathbf{v}',j',\epsilon')$ in the Fock space is defined as

$$A_F = \sum_{j,\epsilon,j',\epsilon'}\int\int A(\mathbf{v},j,\epsilon;\mathbf{v}',j',\epsilon')a(\mathbf{v},j,\epsilon)^*a(\mathbf{v}',j',\epsilon')d\rho(\mathbf{v})d\rho(\mathbf{v}') \tag{3.47}$$

One can directly verify that this is an implementation of IR in the Fock space. In particular, the commutation relations in the Fock space will be preserved regardless of whether the (a,a^*) operators satisfy commutation or anticommutation relations and, if any two operators are adjoint in the implementation of IR described above, they will be adjoint in the Fock space as well. In other words, we have a * homomorphism of Lie algebras of operators acting in the space of IR and in the Fock space.

We now require that in Poincare approximation the energy should be positive definite. Recall that the operators (3.16) and (3.17) act on their respective subspaces or in other words, they are diagonal in the quantum number ϵ.

Suppose that $m_{dS} > 0$ and consider the quantized operator corresponding to the dS energy $\mathcal{E}$ in Eq. (3.16). In Poincare approximation, $\mathcal{E}^{(+)} = m_{dS}v_0$ is fully analogous to the standard free energy and therefore, as follows from Eq. (3.47), its quantized form is

$$(\mathcal{E}^{(+)})_F = m_{dS}\sum_j \int v_0 a(\mathbf{v},j,1)^* a(\mathbf{v},j,1)d\rho(\mathbf{v}) \tag{3.48}$$

This expression is fully analogous to the quantized Hamiltonian in standard theory and it is known that the operator defined in such a way is positive definite.

Consider now the operator $M_{04}^{(-)}$. In Poincare approximation its quantized form is

$$(\mathcal{E}^{(-)})_F == -m_{dS}\sum_j \int v_0 a(\mathbf{v},j,-1)^* a(\mathbf{v},j,-1)d\rho(\mathbf{v}) \tag{3.49}$$

and this operator is negative definite, what is unacceptable.

One might say that the operators $a(\mathbf{v},j,-1)$ and $a(\mathbf{v},j,-1)^*$ are "non-physical": $a(\mathbf{v},j,-1)$ is the operator of object's annihilation with the negative energy, and $a(\mathbf{v},j,-1)^*$ is the operator of object's creation with the negative energy.

We will interpret the operator $(\mathcal{E}^{(-)})_F$ as that related to antiparticles. In QFT the annihilation and creation operators for antiparticles are present in quantized fields with the coefficients describing negative energy solutions of the corresponding covariant equation. This is an implicit implementation of the idea that the creation or annihilation of an antiparticle can be treated, respectively as the annihilation or creation of the corresponding particle with the negative energy. In our case this idea can be implemented explicitly.

Instead of the operators $a(\mathbf{v},j,-1)$ and $a(\mathbf{v},j,-1)^*$, we define new operators $b(\mathbf{v},j)$ and $b(\mathbf{v},j)^*$. If $b(\mathbf{v},j)$ is treated as the "physical" operator of antiparticle annihilation then, according to the above idea, it should be proportional to $a(\mathbf{v},-j,-1)^*$. Analogously, if $b(\mathbf{v},j)^*$ is the "physical" operator of antiparticle creation, it should be proportional to $a(\mathbf{v},-j,-1)$. Therefore

$$b(\mathbf{v},j) = \eta(j)a(\mathbf{v},-j,-1)^* \quad b(\mathbf{v},j)^* = \eta(j)^* a(\mathbf{v},-j,-1) \tag{3.50}$$

where $\eta(j)$ is a phase factor such that

$$\eta(j)\eta(j)^* = 1 \tag{3.51}$$

As follows from this relations, if a particle is characterized by additive quantum numbers (e.g., electric, baryon or lepton charges) then its antiparticle is characterized by the same quantum numbers but with the minus sign. The transformation described by Eqs. (3.50) and (3.51) can also be treated as a special case of the Bogolubov transformation discussed in a wide literature on many-body theory (see, e.g., Chap. 10 in Ref. [88] and references therein).

Since we treat $b(\mathbf{v},j)$ as the annihilation operator and $b(\mathbf{v},j)^*$ as the creation one, instead of Eq. (3.45) we should define a new vacuum state $\tilde{\Phi}_0$ such that

$$a(\mathbf{v},j,1)\tilde{\Phi}_0 = b(\mathbf{v},j)\tilde{\Phi}_0 = 0 \quad \forall\ \mathbf{v},j, \tag{3.52}$$

and the images of states localized in X_- should be defined as

$$\varphi_F^{(-)} = \sum_{j,\epsilon} \int \varphi(\mathbf{v},j,-1) b(\mathbf{v},j)^* d\rho(\mathbf{v}) \tilde{\Phi}_0 \tag{3.53}$$

In that case the (b,b^*) operators should be such that in the case of anticommutation relations

$$\{b(\mathbf{v},j), b(\mathbf{v}',j')^*\} = \delta_{jj'} v_0 \delta^{(3)}(\mathbf{v}-\mathbf{v}'), \tag{3.54}$$

and in the case of commutation relations

$$[b(\mathbf{v},j), b(\mathbf{v}',j')^*] = \delta_{jj'} v_0 \delta^{(3)}(\mathbf{v}-\mathbf{v}') \tag{3.55}$$

We have to verify whether the new definition of the vacuum and one-particle states is a correct implementation of IR in the Fock space. A necessary condition is that the new operators should satisfy the commutation relations of the dS algebra. Since we replaced the (a,a^*) operators by the (b,b^*) operators only if $\epsilon=-1$, it is obvious from Eq. (3.47) that the images of the operators (3.16) in the Fock space satisfy Eq. (1.4). Therefore we have to verify that the images of the operators (3.17) in the Fock space also satisfy Eq. (1.4).

Consider first the case when the operators $a(\mathbf{v},j,\epsilon)$ satisfy the anticommutation relations. By using Eq. (3.50) one can express the operators $a(\mathbf{v},j,-1)$ in terms of the operators $b(\mathbf{v},j)$. Then it follows from the condition (3.50) that the operators $b(\mathbf{v},j)$ indeed satisfy Eq. (3.55). If the operator A_F is defined by Eq. (3.47) and is expressed only in terms of the (a,a^*) operators at $\epsilon=-1$, then in terms of the (b,b^*)-operators it acts on states localized in X_- as

$$A_F = \sum_{j,j'} \int\int A(\mathbf{v},j,-1;\mathbf{v}',j',-1)\eta(j')\eta(j)^* b(\mathbf{v},-j) b(\mathbf{v}',-j')^* d\rho(\mathbf{v}) d\rho(\mathbf{v}') \tag{3.56}$$

As follows from Eq. (3.55), this operator can be written as

$$A_F = C - \sum_{j,j'} \int\int A(\mathbf{v}',-j',-1;\mathbf{v},-j,-1)\eta(j)\eta(j')^* b(\mathbf{v},j)^* b(\mathbf{v}',j') d\rho(\mathbf{v}) d\rho(\mathbf{v}') \tag{3.57}$$

where C is the trace of the operator A_F

$$C = \sum_j \int A(\mathbf{v}, j, -1; \mathbf{v}, j, -1) d\rho(\mathbf{v}) \tag{3.58}$$

and in general it is some an indefinite constant. The existence of infinities in the standard approach is the well-known problem. Usually the infinite constant is eliminated by requiring that all quantized operators should be written in the normal form or by using another prescriptions. However, in dS theory this constant cannot be eliminated since IRs are defined on the space which is a direct some of X_+ and X_-, and the constant inevitably arise when one wishes to have an interpretation of IRs in terms of particles and antiparticles. In Sec. 8.8 we consider an example when a constant, which is infinite in standard theory, becomes zero in GFQT but this result can be obtained only if the IR is implemented by using a basis characterized by discrete numbers.

In this chapter we assume that neglecting the constant C can be somehow justified. In that case if the operator A_F is defined by Eq. (3.47) then in the case of anticommutation relations its action on states localized in X_- can be written as in Eq. (3.57) with $C = 0$. Then, taking into account the properties of the kernels discussed above, we conclude that in terms of the (b, b^*)-operators the kernels of the operators $(m_{dS}v)_F$ change their sign while the kernels of the operators in Eq. (3.41) remain the same. In particular, the operator $(-m_{dS}v_0)_F$ acting on states localized on X_- has the same kernel as the operator $(m_{ds}v_0)_F$ acting on states localized in X_+ has in terms of the a-operators. This implies that in Poincare approximation the energy of the states localized in X_- is positive definite, as well as the energy of the states localized in X_+.

Consider now how the spin operator changes when the a-operators are replaced by the b-operators. Since the spin operator is diagonal in the variable $\mathbf{v}$, it follows from Eq. (3.57) that the transformed spin operator will have the same kernel if

$$s_i(j, j') = -\eta(j)\eta(j')^* s_i(-j', -j) \tag{3.59}$$

where $s_i(j, j')$ is the kernel of the operator s_i. For the z component of the spin operator this relation is obvious since s_z is diagonal in (j, j') and its kernel is $s_z(j, j') = j\delta_{jj'}$.If we choose $\eta(j) = (-1)^{(s-j)}$ then the validity of Eq. (3.59) for $s = 1/2$ can be verified directly while in the general case it can be verified by using properties of $3j$ symbols.

The above results for the case of anticommutation relations can be summarized as follows. If we replace m_{dS} by $-m_{dS}$ in Eq. (3.17) then the new set of operators

$$\mathbf{J}' = l(\mathbf{v}) + \mathbf{s}, \quad \mathbf{N}' = -iv_0\frac{\partial}{\partial \mathbf{v}} + \frac{\mathbf{s}\times\mathbf{v}}{v_0+1},$$
$$\mathbf{B}' = m_{dS}\mathbf{v} - i[\frac{\partial}{\partial \mathbf{v}} + \mathbf{v}(\mathbf{v}\frac{\partial}{\partial \mathbf{v}}) + \frac{3}{2}\mathbf{v}] - \frac{\mathbf{s}\times\mathbf{v}}{v_0+1},$$
$$\mathcal{E}' = m_{dS}v_0 - iv_0(\mathbf{v}\frac{\partial}{\partial \mathbf{v}} + \frac{3}{2}) \tag{3.60}$$

obviously satisfies the commutation relations (1.4). The kernels of these operators define quantized operators in terms of the (b, b^*)-operators in the same way as the kernels of the operators (3.16) define quantized operators in terms of the (a, a^*)-operators. In particular, in Poincare approximation the energy operator acting on states localized in X_- can be defined as $E' = \mathcal{E}'/R$ and in this approximation it is positive definite.

At the same time, in the case of commutation relation the replacement of the (a, a^*)-operators by the (b, b^*)-operators is unacceptable for several reasons. First of all, if the operators $a(\mathbf{v}, j, \epsilon)$ satisfy the commutation relations (3.44), the operators defined by Eq. (3.50) will not satisfy Eq. (3.55). Also, the r.h.s. of Eq. (3.57) will now have the opposite sign. As a result, the transformed operator $\mathcal{E}$ will remain negative definite in Poincare approximation and the operators (3.41) will change their sign. In particular, the angular momentum operators will no longer satisfy correct commutation relations.

We have shown that if the definitions (3.45) and (3.46) are replaced by (3.52) and (3.53), respectively, then the images of both sets of operators in Eq. (3.16) and Eq. (3.17) satisfy the correct commutation relations in the case of anticommutators. A question arises whether the new implementation in the Fock space is equivalent to the IR described in Sec. 3.2. For understanding the essence of the problem, the following very simple pedagogical example might be useful.

Consider a representation of the SO(2) group in the space of functions $f(\varphi)$ on the circumference $\varphi \in [0, 2\pi]$ where φ is the polar angle and the points $\varphi = 0$ and $\varphi = 2\pi$ are identified. The generator of counterclockwise rotations is $A = -id/d\varphi$ while the generator of clockwise rotations is $B = id/d\varphi$. The equator of the circumference contains two points, $\varphi = 0$ and $\varphi = \pi$ and has measure zero. Therefore we can represent each $f(\varphi)$ as a superposition of functions with the carriers in the upper and lower semi circumferences, S_+ and S_-. The operators A and B are defined only on differentiable functions. The Hilbert space H contains not only such functions but a set of differentiable functions is dense in H. If a function $f(\varphi)$ is differentiable and has the carrier in S_+ then $Af(\varphi)$ and $Bf(\varphi)$ also have the carrier in S_+ and analogously for functions with the carrier in S_-. However, we cannot define a representation of the SO(2) group such that its generator is A on functions with the carrier in S_+ and B on functions with the carrier in S_- because a counterclockwise rotation on S_+ should be counterclockwise on S_- and analogously for clockwise rotations. In other words, the actions of the generator on functions with the carriers in S_+ and S_- cannot be independent.

In the case of finite dimensional representations, any IR of a Lie algebra by Hermitian operators can be always extended to an UIR of the corresponding Lie group. In that case the UIR has a property that any state is its cyclic vector i.e. the whole representation space can be obtained by acting by representation operators on this vector and taking all possible linear combinations. For infinite dimensional IRs this is not always the case and there should exist conditions for IRs of Lie algebras by Hermitian operators to be extended to corresponding UIRs. This problem has been extensively discussed in the mathematical literature (see e.g. Ref. [82]). By analogy with finite dimensional IRs, one might think that in the case of infinite dimensional IRs there should exist an analog of the cyclic vector. In Sec. 3.4 we have shown that for infinite dimensional IRs of the dS algebra this idea can be explicitly implemented by choosing a cyclic vector and acting on this vector by operators of the enveloping algebra of the dS algebra. This construction shows that the action of representation operators on states with the carrier in X_+ should define its action on states with the carrier in X_-, i.e. the action of representation operators on states with the carriers in X_+ and X_- are not independent.

3.5.2 Example of transformation mixing particles and antiparticles

We treated states localized in X_+ as particles and states localized in X_- as corresponding antiparticles. However, the space of IR contains not only such states. There is no rule prohibiting states with the carrier having a nonempty intersection with both, X_+ and X_-. Suppose that there exists a unitary transformation belonging to the UIR of the dS group such that it transforms a state with the carrier in X_+ to a state with the carrier in X_-. If the Fock space is implemented according to Eqs. (3.45) and (3.46) then the transformed state will have the form

$$\varphi_F^{(-)} = \sum_j \int \varphi(\mathbf{v}, j) a(\mathbf{v}, j, -1)^* d\rho(\mathbf{v}) \Phi_0 \tag{3.61}$$

while with the implementation in terms of the (b, b^*) operators it should have the form (3.53). Since the both states are obtained from the same state with the carrier in X_+, they should be the same. However, they cannot be the same. This is clear even from the fact that in Poincare approximation the former has a negative energy while the latter has a positive energy.

Our construction shows that the interpretation of states as particles and antiparticles is not always consistent. It can be only approximately consistent when we consider only states localized either in X_+ or in X_- and only transformations which do not mix such states. In quantum theory there is a superselection rule (SSR) prohibiting states which are superpositions of states with different electric, baryon or lepton charges. In general, if states ψ_1 and ψ_2 are such that there are no physical operators A such that $(\psi_2, A\psi_1) \neq 0$ then the SSR says that the state $\psi = \psi_1 + \psi_2$ is prohibited. The meaning of the SSR is now widely discussed (see e.g., Ref. [89] and references therein). Since the SSR implies that the superposition principle, which is a key principle of quantum theory, is not universal, several authors argue that the SSR should not be present in quantum theory. Other authors argue that the SSR is only a dynamical principle since, as a result of decoherence, the state ψ will quickly disappear and so it cannot be observable.

We now give an example of a transformation, which transforms states localized in X_+ to ones localized in X_- and *vice versa*. Let $I \in SO(1,4)$ be a matrix which formally coincides with the metric tensor η. If this matrix is treated as a transformation of the dS space, it transforms the North pole $(0,0,0,0,x^4 = R)$ to the South pole $(0,0,0,0,x^4 = -R)$ and *vice versa*. As already explained, in our approach the dS space is not involved and in Secs. 3.2–3.4 the results for UIRs of the dS group have been used only for constructing IRs of the dS algebra. This means that the unitary operator $U(I)$ corresponding to I is well defined and we can consider its action without relating I to a transformation of the dS space.

If $\mathbf{v}_L$ is a representative defined by Eq. (3.13) then it is easy to verify that $I\mathbf{v}_L = (-\mathbf{v})_L I$ and, as follows from Eq. (3.9), if ψ_1 is localized in X_+ then $\psi_2 = U(I)\psi_1$ will be localized in X_-. Therefore $U(I)$ transforms particles into antiparticles and *vice versa*. In Secs. 1.2 and 1.3 we argued that the notion of empty space-time background is unphysical and that unitary transformations generated by self-adjoint operators may not have a usual interpretation. The example with $U(I)$ gives a good illustration of this point. Indeed, if we work with dS space, we might expect that all unitary transformations corresponding to the elements of the group SO(1,4) act in the space of IR only kinematically, in particular they transform particles to particles and antiparticles to antiparticles. However, in QFT in curved space-time this is not the case. Nevertheless, this is not treated as an indication that standard notion of the dS space is not physical. Although fields are not observable, in QFT in curved space-time they are treated as fundamental and single-particle interpretations of field equations are not tenable (moreover, some QFT theorists state that particles do not exist). For example, as shown in Ref. [90], solutions of fields equations are superpositions of states which usually are interpreted as a particle and its antiparticle, and in dS space neither coefficient in the superposition can be zero. This result is compatible with the Mensky's one [46] described in the beginning of this section. One might say that our result is in agreement with those in dS QFT since UIRs of the dS group describe not a particle or antiparticle but an object such that a particle and its antiparticle are different states of this object (at least in Poincare approximation). However, as noted above, in dS QFT this is not treated as the fact that dS space is unphysical.

The matrix I belongs to the component of unity of the group SO(1,4). For example, the transformation I can be obtained as a product of rotations by 180 degrees in planes $(1,2)$ and $(3,4)$. Therefore, $U(I)$ can be obtained as a result of continuous transformations $exp[i(M_{12}\varphi_1 + M_{34}\varphi_2)]$ when the values of φ_1 and φ_2 change from zero to π. Any continuous transformation transforming a state with the carrier in X_+ to the state with the carrier in X_- is such that the carrier should cross X_0 at some values of the transformation parameters. As noted in the preceding section, the set X_0 is characterized by the condition that the standard Poincare momentum is infinite and therefore, from the point of view of intuition based on Poincare invariant theory, one might think that no transformation when the carrier crosses X_0 is possible. However, as we have seen in the preceding section, in variables (u_1, u_2, u_3, u_4) the condition $u_4 = 0$ defines the equator of S^3 corresponding to X_0 and this condition is not singular. So from the point of view of dS theory, nothing special happens when the carrier crosses X_0. We observe only either particles or antiparticles but not their linear combinations because Poincare approximation works with a very high accuracy and it is very difficult to perform transformations mixing states localized in X_+ and X_-.

3.5.3 Summary

As follows from the above discussion, *objects belonging to IRs of the dS algebra can be treated as particles or antiparticles only if Poincare approximation works with a high accuracy.* As a consequence, *the conservation of electric, baryon and lepton charges can be only approximate.*

At the same time, our discussion shows that the approximation when one IR of the dS algebra splits into independent IRs for a particle and its antiparticle can be valid only in the case of anticommutation relations. Since it is a reasonable requirement that dS theory should become the Poincare one at certain conditions, the above results show that *in dS invariant theory only fermions can be elementary.*

Let us now consider whether there exist neutral particles in dS invariant theory. In AdS and Poincare invariant theories, neutral particles are described as follows. One first constructs a covariant field containing both IRs, with positive and negative energies. Therefore the number of states is doubled in comparison with the IR. However, to satisfy the requirement that neutral particles should be described by real (not complex) fields, one has to impose a relation between the creation and annihilation operators for states with positive and negative energies. Then the number of states describing a neutral field again becomes equal to the number of states in the IR. In contrast to those theories, IRs of the dS algebra are implemented on both, upper and lower Lorentz hyperboloids and therefore the number of states in IRs is twice as big as for IRs of the Poincare and AdS algebras. Even this fact shows that in dS invariant theory there can be no neutral particles since it is not possible to reduce the number of states in an IR. Another argument is that, as follows from the above construction, dS invariant theory is not C invariant. Indeed, C invariance in standard theory means that representation operators are invariant under the interchange of a-operators and b-operators. However, in our case when a-operators are replaced by b-operators, the operators (3.16) become the operators (3.60). Those sets of operators coincide only in Poincare approximation while in general the operators $M^{4\mu}$ in Eqs. (3.16) and (3.60) are different. Therefore a particle and its antiparticle are described by different sets of operators. We conclude that *in dS invariant theory neutral particles cannot be elementary.*

3.6 dS quantum mechanics and cosmological repulsion

The results on IRs can be applied not only to elementary particles but even to macroscopic bodies when it suffices to consider their motion as a whole. This is the case when the distances between the bodies are much greater that their sizes. In this section we consider the operators $M^{4\mu}$ not only in Poincare approximation but taking into account dS corrections. If those corrections are small, one can neglect transformations mixing states on the upper and lower Lorentz hyperboloids (see the discussion in the preceding section) and describe the representation operators for a particle and its antiparticle by Eqs. (3.16) and (3.60), respectively.

We define $E = \mathcal{E}/R$, $\mathbf{P} = \mathbf{B}/R$ and $m = m_{dS}/R$. Consider the nonrelativistic approximation when $|\mathbf{v}| \ll 1$. If we wish to work with units where the dimension of velocity is m/s, we should replace $\mathbf{v}$ by $\mathbf{v}/c$. If $\mathbf{p} = m\mathbf{v}$ then it is clear from the expressions for $\mathbf{B}$ in Eqs. (3.16) and (3.60) that $\mathbf{p}$ becomes the real momentum $\mathbf{P}$ only in the limit $R \to \infty$. At this stage we do not have any coordinate space yet. However, if we assume that semiclassical approximation is valid, then, by analogy with standard quantum mechanics, we can *define* the position operator $\mathbf{r}$ as $i\partial/\partial\mathbf{p}$. As discussed in Chap. 2, such a definition encounters problems in view of the WPS effect. However, as noted in this chapter, this effect is a pure quantum phenomenon and for macroscopic bodies it is negligible. The problem of the cosmological acceleration is meaningful only for macroscopic bodies when classical approximation applies.

Since the commutators of $\mathcal{R}_{||}$ and $\mathcal{R}_{\perp}$ with different components of $\mathbf{p}$ are proportional to $\hbar$ and the operator $\mathbf{r}$ is a sum of the parallel and perpendicular components (see Eq. (9.6)), in classical approximation we can neglect those commutators and treat $\mathbf{p}$ and $\mathbf{r}$ as usual vectors. Then as follows from Eq. (3.16)

$$\mathbf{P} = \mathbf{p} + mc\mathbf{r}/R, \quad H = \mathbf{p}^2/2m + c\mathbf{p}\mathbf{r}/R, \quad \mathbf{N} = -m\mathbf{r} \tag{3.62}$$

where $H = E - mc^2$ is the classical nonrelativistic Hamiltonian and, as follows from Eqs. (3.60)

$$\mathbf{P} = \mathbf{p} - mc\mathbf{r}/R, \quad H = \mathbf{p}^2/2m - c\mathbf{p}\mathbf{r}/R, \quad \mathbf{N} = -m\mathbf{r} \tag{3.63}$$

As follows from these expressions, in both cases

$$H(\mathbf{P},\mathbf{r}) = \frac{\mathbf{P}^2}{2m} - \frac{mc^2\mathbf{r}^2}{2R^2} \tag{3.64}$$

The last term in Eq. (3.64) is the dS correction to the non-relativistic Hamiltonian. It is interesting to note that the non-relativistic Hamiltonian depends on c although it is usually believed that c can be present only in relativistic theory. This illustrates the fact mentioned in Sec. 1.4 that the transition to nonrelativistic theory understood as $|\mathbf{v}| \ll 1$ is more physical than that understood as $c \to \infty$. The presence of c in Eq. (3.64) is a consequence of the fact that this expression is written in standard units. In nonrelativistic theory c is usually treated as a very large quantity. Nevertheless, the last term in Eq. (3.64) is not large since we assume that R is very large.

The result given by Eq. (1.7) is now a consequence of the Hamilton equations for the Hamiltonian given by Eq. (3.64). In our approach this result has been obtained without using dS space and Riemannian geometry while the fact that $\Lambda \neq 0$ should be treated not such that the space-time background has a curvature (since the notion of the space-time background is meaningless) but as an indication that the symmetry algebra is the dS algebra rather than the Poincare one. *Therefore for explaining the fact that $\Lambda \neq 0$ there is no need to involve dark energy or any other quantum fields.*

Another way to show that our results are compatible with GR is as follows. The well-known result of GR is that if the metric is stationary and differs slightly from the Minkowskian one then in the nonrelativistic approximation the curved space-time can be effectively described by a gravitational potential $\varphi(\mathbf{r}) = (g_{00}(\mathbf{r}) - 1)/2c^2$. We now express x_0 in Eq. (1.5) in terms of a new variable t as $x_0 = t + t^3/6R^2 - t\mathbf{x}^2/2R^2$. Then the expression for the interval becomes

$$ds^2 = dt^2(1 - \mathbf{r}^2/R^2) - d\mathbf{r}^2 - (\mathbf{r}d\mathbf{r}/R)^2 \tag{3.65}$$

Therefore, the metric becomes stationary and $\varphi(\mathbf{r}) = -\mathbf{r}^2/2R^2$ in agreement with Eq. (3.64).

Consider now a system of two free particles described by the variables $\mathbf{P}_j$ and $\mathbf{r}_j$ $(j = 1,2)$. Define the standard nonrelativistic variables

$$\begin{aligned} &\mathbf{P}_{12} = \mathbf{P}_1 + \mathbf{P}_2, \quad \mathbf{q}_{12} = (m_2\mathbf{P}_1 - m_1\mathbf{P}_2)/(m_1 + m_2) \\ &\mathbf{R}_{12} = (m_1\mathbf{r}_1 + m_2\mathbf{r}_2)/(m_1 + m_2), \quad \mathbf{r}_{12} = \mathbf{r}_1 - \mathbf{r}_2 \end{aligned} \tag{3.66}$$

Then, as follows from Eqs. (3.62) and (3.63), in the nonrelativistic approximation the two-particle quantities $\mathbf{P}$, $\mathbf{E}$ and $\mathbf{N}$ are given by

$$\mathbf{P} = \mathbf{P}_{12}, \quad E = M + \frac{\mathbf{P}_{12}^2}{2M} - \frac{Mc^2\mathbf{R}_{12}^2}{2R^2}, \quad \mathbf{N} = -M\mathbf{R}_{12} \tag{3.67}$$

where

$$M = M(\mathbf{q}_{12}, \mathbf{r}_{12}) = m_1 + m_2 + H_{nr}(\mathbf{r}_{12}, \mathbf{q}_{12}), \quad H_{nr}(\mathbf{r}, \mathbf{q}) = \frac{\mathbf{q}^2}{2m_{12}} - \frac{m_{12}c^2\mathbf{r}^2}{2R^2} \quad (3.68)$$

and m_{12} is the reduced two-particle mass.

It now follows from Eqs. (3.18) and (3.67) that M has the meaning of the two-body mass since in the nonrelativistic approximation $M^2 = I_2/R^2$ where now I_2 is the Casimir operator of the second order for the two-body system. Therefore $M(\mathbf{q}_{12}, \mathbf{r}_{12})$ is the internal two-body Hamiltonian. Then, as a consequence of the Hamilton equations, in semiclassical approximation the relative acceleration is given by the same expression (1.7) but now $\mathbf{a}$ is the relative acceleration and $\mathbf{r}$ is the relative radius vector. As noted in Sec. 1.2, equations of motions for systems of free particles can be obtained even without the Hamilton equations but assuming that the coordinates and momenta are related to each other by Eq. (1.2). This question is discussed in Sec. 5.7.

The fact that two free particles have a relative acceleration is known for cosmologists who consider dS symmetry on classical level. This effect is called the dS antigravity. The term antigravity in this context means that the particles repulse rather than attract each other. In the case of the dS antigravity the relative acceleration of two free particles is proportional (not inversely proportional!) to the distance between them. This classical result (which in our approach has been obtained without involving dS space and Riemannian geometry) is a special case of dS symmetry on quantum level when semiclassical approximation works with a good accuracy.

As follows from Eq. (3.68), the dS antigravity is not important for local physics when $r \ll R$. At the same time, at cosmological distances the dS antigravity is much stronger than any other interaction (gravitational, electromagnetic *etc.*). One can consider the quantum two-body problem with the Hamiltonian given by Eq. (3.68). Then it is obvious that the spectrum of the operator H_{nr} is purely continuous and belongs to the interval $(-\infty, \infty)$ (see also Refs. [36, 37] for details). This does not mean that the theory is unphysical since stationary bound states in standard theory become quasistationary with a very large lifetime if R is large.

Our final remark follows. The consideration in this chapter involves only standard quantum-mechanical notions and in semiclassical approximation the results on the cosmological acceleration are compatible with GR. As argued in Sect. 2.2, the standard coordinate operator has some properties which do not correspond to what is expected from physical intuition; however, at least from mathematical point of view, at cosmological distances semiclassical approximation is valid with a very high accuracy. At the same time, as discussed in the next chapter, when distances are much less than cosmological ones, this operator should be modified. Then, as a consequence of the fact that in dS invariant theory the spectum of the mass operator for a free two-body system is not bounded below by $(m_1 + m_2)$ it is possible to obtain gravity as a pure kinematical consequence of dS symmetry on quantum level.

Chapter 4

Algebraic description of irreducible representations

4.1 Construction of IRs in discrete basis

In Sec. 3.4 we have mentioned a possibility that IRs of the so(1,4) algebra can be constructed in a pure algebraic approach such that the basis is characterized only by discrete quantum numbers. In this chapter a detailed consideration of this approach is given for the spinless case and in the next chapter the results are applied to gravity. First of all, to make relations between standard theory and GFQT more straightforward, we will modify the commutation relations (1.4) by writing them in the form

$$[M^{ab}, M^{cd}] = -2i(\eta^{ac}M^{bd} + \eta^{bd}M^{ac} - \eta^{ad}M^{bc} - \eta^{bc}M^{ad}) \quad (4.1)$$

One might say that these relations are written in units $\hbar/2 = c = 1$. However, as noted in Sect. 1.4, fundamental quantum theory should not involve quantities $\hbar$ and c at all, and Eq. (4.1) indeed does not contain these quantities. The reason for writing the commutation relations in the form (4.1) rather than (1.4) is that in this case the minimum nonzero value of the angular momentum is 1 instead of 1/2. Therefore the spin of fermions is odd and the spin of bosons is even. This will be convenient in GFQT where 1/2 is a very large number (see Chap. 6).

As already noted, the results on IRs can be applied not only to elementary particles but even to macroscopic bodies when it suffices to consider their motion as a whole. This is the case when the distances between the bodies are much greater that their sizes. In Poincare invariant theory, IRs describing massless Weyl particles can be obtained as a limit of massive IRs when $m \to 0$ with a special choice of representatives in the factor space $SL(2, C)/SU(2)$. However, as shown in Sec. 3.3, in dS theory such a limit does not exist and therefore there are no Weyl particles in dS theory. In standard theory it is believed that the photon is a true massless particle but, as noted in Sec. 3.3, if, for example, R is of the order of $10^{26}m$ then the upper limit for the photon dS mass is of the order of 10^{15}. In the present work we assume that the photon can be described by IRs of the principle series discussed above.

In all macroscopic experiments the orbital angular momenta of macroscopic bodies and even photons are very large. As an example, consider a photon moving in approximately radial direction away from the Earth surface. Suppose that the photon energy equals the bound energy of the ground state of the hydrogen atom 27.2ev. Then in units $c = \hbar = 1$ this energy is of the order of $10^7/cm$. Hence even if the level arm of the photon trajectory is of the order of $1cm$, the value of the orbital angular momentum is of the order of 10^7. In other experiments with photons and macroscopic bodies this value is greater by many orders of magnitude. Therefore the spin terms in $\mathbf{J}$ can be neglected. Since $v_0 > |\mathbf{v}|$, the orbital part of the operator $\mathbf{N}$ is also much greater than its spin part. The orbital part of the operator $\mathbf{B}$ is typically much greater than its spin part; this is clear even from the fact that in Poincare limit this part is proportional to R while the spin does not depend on R. In view of these remarks, we will not consider spin effects. Hence our goal is to construct massive spinless IRs in a discrete basis. By analogy with the method of little group in standard theory, one can first choose states which can be treated as rest ones and then obtain the whole representation space by acting on such states by certain linear combinations of representation operators.

Since $\mathbf{B}$ is a possible choice of the dS analog of the momentum operator, one might think that rest states e_0 can be defined by the condition $\mathbf{B}e_0 = 0$. However, in the general case this is not consistent since, as follows from Eq. (4.1), different components of $\mathbf{B}$ do not commute with each other: as follows from Eq. (4.1) and the definitions of the operators $\mathbf{J}$ and $\mathbf{B}$ in Sect. 3.2,

$$[J^j, J^k] = [B^j, B^k] = 2ie_{jkl}J^l, \quad [J^j, B^k] = 2ie_{jkl}B^l \tag{4.2}$$

where a sum over repeated indices is assumed. Therefore a subspace of elements e_0 such that $B^j e_0 = 0$ $(j = 1, 2, 3)$ is not closed under the action of the operators B^j.

Let us define the operators $\mathbf{J}' = (\mathbf{J}+\mathbf{B})/2$ and $\mathbf{J}" = (\mathbf{J}-\mathbf{B})/2$. As follows from Eq. (4.1), they satisfy the commutation relations

$$[J'^j, J"^k] = 0, \quad [J'^j, J'^k] = 2ie_{jkl}J'^l, \quad [J"^j, J"^k] = 2ie_{jkl}J"^l \tag{4.3}$$

Since in Poincare limit $\mathbf{B}$ is much greater than $\mathbf{J}$, as an analog of the momentum operator one can treat $\mathbf{J}'$ instead of $\mathbf{B}$. Then one can define rest states e_0 by the condition that $\mathbf{J}'e_0 = 0$. In this case the subspace of rest states is defined consistently since it is invariant under the action of the operators $\mathbf{J}'$. Since the operators $\mathbf{J}'$ and $\mathbf{J}"$ commute with each other, one can define the internal angular momentum of the system as a reduction of $\mathbf{J}"$ on the subspace of rest states. In particular, in Ref. [35] we used such a construction for constructing IRs of the dS algebra in the method of $SU(2)\times SU(2)$ shift operators proposed by Hughes for constructing IRs of the SO(5) group [87]. In the spinless case the situation is simpler since for constructing IRs it suffices to choose only one vector e_0 such that

$$\mathbf{J}'e_0 = \mathbf{J}"e_0 = 0, \quad I_2 e_0 = (w+9)e_0 \tag{4.4}$$

The last requirement reflects the fact that all elements from the representation space are eigenvectors of the Casimir operator I_2 with the same eigenvalue. When the representation operators satisfy Eq. (4.1), the numerical value of the operator I_2 is not as indicated at the end of Sec. (3.2) but

$$I_2 = w - s(s+2) + 9 \tag{4.5}$$

where $w = m_{dS}^2$. Therefore for spinless particles the numerical value equals $w + 9$.

As follows from Eq. (4.1) and the definitions of the operators $(\mathbf{J}, \mathbf{N}, \mathbf{B}, \mathcal{E})$ in Secs. 3.2 and (3.4), in addition to Eqs. 4.2, the following relations are satisfied:

$$[\mathcal{E}, \mathbf{N}] = 2i\mathbf{B}, \ [\mathcal{E}, \mathbf{B}] = 2i\mathbf{N}, \ [\mathbf{J}, \mathcal{E}] = 0, \ [B^j, N^k] = 2i\delta_{jk}\mathcal{E}, \ [J^j, N^k] = 2ie_{jkl}N^l \tag{4.6}$$

We define $e_1 = 2\mathcal{E}e_0$ and

$$e_{n+1} = 2\mathcal{E}e_n - [w + (2n+1)^2]e_{n-1} \tag{4.7}$$

These definitions make it possible to find e_n for any $n = 0, 1, 2....$ As follows from Eqs. (4.2), (4.6) and (4.7), $\mathbf{J}e_n = 0$ and $\mathbf{B}^2 e_n = 4n(n+2)e_n$. We use the notation $J_x = J^1$, $J_y = J^2$, $J_z = J^3$ and analogously for the operators $\mathbf{N}$ and $\mathbf{B}$. Instead of the (xy) components of the vectors it may be sometimes convenient to use the $\pm$ components such that $J_x = J_+ + J_-$, $J_y = -i(J_+ - J_-)$ and analogously for the operators $\mathbf{N}$ and $\mathbf{B}$. We now define the elements e_{nkl} as

$$e_{nkl} = \frac{(2k+1)!!}{k!l!}(J_-)^l(B_+)^k e_n \tag{4.8}$$

Then a direct calculation using Eqs. (4.2-4.8) gives

$$
\begin{aligned}
&\mathcal{E}e_{nkl} = \frac{n+1-k}{2(n+1)}e_{n+1,kl} + \frac{n+1+k}{2(n+1)}[w+(2n+1)^2]e_{n-1,kl}\\
&N_+e_{nkl} = \frac{i(2k+1-l)(2k+2-l)}{8(n+1)(2k+1)(2k+3)}\{e_{n+1,k+1,l} -\\
&[w+(2n+1)^2]e_{n-1,k+1,l}\} -\\
&\frac{i}{2(n+1)}\{(n+1-k)(n+2-k)e_{n+1,k-1,l-2} -\\
&(n+k)(n+1+k)[w+(2n+1)^2]e_{n-1,k-1,l-2}\}\\
&N_-e_{nkl} = \frac{-i(l+1)(l+2)}{8(n+1)(2k+1)(2k+3)}\{e_{n+1,k+1,l+2} -\\
&[w+(2n+1)^2]e_{n-1,k+1,l+2}\} +\\
&\frac{i}{2(n+1)}\{(n+1-k)(n+2-k)e_{n+1,k-1,l} -\\
&(n+k)(n+1+k)[w+(2n+1)^2]e_{n-1,k-1,l}\}\\
&N_ze_{nkl} = \frac{-i(l+1)(2k+1-l)}{4(n+1)(2k+1)(2k+3)}\{e_{n+1,k+1,l+1} -\\
&[w+(2n+1)^2]e_{n-1,k+1,l+1}\} -\\
&\frac{i}{n+1}\{(n+1-k)(n+2-k)e_{n+1,k-1,l-1} -\\
&(n+k)(n+1+k)[w+(2n+1)^2]e_{n-1,k-1,l-1}\}
\end{aligned}
\qquad (4.9)
$$

$$
\begin{aligned}
&B_+e_{nkl} = \frac{(2k+1-l)(2k+2-l)}{2(2k+1)(2k+3)}e_{n,k+1,l} -\\
&2(n+1-k)(n+1+k)e_{n,k-1,l-2}\\
&B_-e_{nkl} = \frac{(l+1)(l+2)}{2(2k+1)(2k+3)}e_{n,k+1,l+2} +\\
&2(n+1-k)(n+1+k)e_{n,k-1,l}\\
&B_ze_{nkl} = \frac{(l+1)(2k+1-l)}{2(2k+1)(2k+3)}e_{n,k+1,l+1} -\\
&4(n+1-k)(n+1+k)e_{n,k-1,l-1}\\
&J_+e_{nkl} = (2k+1-l)e_{nk,l-1} \quad J_-e_{nkl} = (l+1)e_{nk,l+1}\\
&J_ze_{nkl} = 2(k-l)e_{nkl}
\end{aligned}
\qquad (4.10)
$$

where at a fixed value of n, $k = 0,1,...n$, $l = 0,1,...2k$ and if l and k are not in this range then $e_{nkl} = 0$. Therefore, the elements e_{nkl} form a basis of the spinless IR with a given w.

The next step is to define a scalar product compatible with the Hermiticity of the operators $(\mathcal{E}, \mathbf{B}, \mathbf{N}, \mathbf{J})$. Since $\mathbf{B}^2 + \mathbf{J}^2$ is the Casimir operator for the so(4) subalgebra and

$$(\mathbf{B}^2 + \mathbf{J}^2)e_{nkl} = 4n(n+2)e_{nkl} \tag{4.11}$$

the vectors e_{nkl} with different values of n should be orthogonal. Since $\mathbf{J}^2$ is the Casimir operator of the so(3) subalgebra and $\mathbf{J}^2 e_{nkl} = 4k(k+1)e_{nkl}$, the vectors e_{nkl} with different values of k also should be orthogonal. Finally, as follows from the last expression in Eq. (4.10), the vectors e_{nkl} with the same values of n and k and different values of l should be orthogonal since they are eigenvectors of the operator J_z with different eigenvalues. Therefore, the scalar product can be defined assuming that $(e_0, e_0) = 1$ and a direct calculation using Eqs. (4.4-4.8) gives

$$(e_{nkl}, e_{nkl}) = (2k+1)! C_{2k}^l C_n^k C_{n+k+1}^k \prod_{j=1}^{n} [w + (2j+1)^2] \tag{4.12}$$

where $C_n^k = n!/[(n-k)!k!]$ is the binomial coefficient. At this point we do not normalize basis vectors to one since, as will be discussed below, the normalization (4.12) has its own advantages.

Instead of l we define a new quantum number $\mu = k - l$ which can take values $-k, -k+1, ...k$. Each element of the representation space can be written as $x = \sum_{nk\mu} c(n,k,\mu)e_{nk\mu}$ where the set of the coefficients $c(n,k,\mu)$ can be called the wave function in the $(nk\mu)$ representation. As follows from Eqs. (4.9) and (4.10), the action of the representation operators on the wave function can be written as

$$\mathcal{E}c(n,k,\mu) = \frac{n-k}{2n}c(n-1,k,\mu) + \frac{n+2+k}{2(n+2)}[w+(2n+3)^2]$$
$$c(n+1,k,\mu)$$
$$N_+c(n,k,\mu) = \frac{i(k+\mu)(k+\mu-1)}{8(2k-1)(2k+1)}\{\frac{1}{n}c(n-1,k-1,\mu-1) -$$
$$\frac{1}{n+2}[w+(2n+3)^2]c(n+1,k-1,\mu-1)\} -$$
$$\frac{i(n-1-k)(n-k)}{2n}c(n-1,k+1,\mu-1) +$$
$$\frac{i(n+k+2)(n+k+3)}{2(n+2)}[w+(2n+3)^2]c(n+1,k+1,\mu-1)$$
$$N_-c(n,k,\mu) = \frac{-i(k-\mu)(k-\mu-1)}{8(2k-1)(2k+1)}\{\frac{1}{n}c(n-1,k-1,\mu+1) -$$
$$\frac{1}{n+2}[w+(2n+3)^2]c(n+1,k-1,\mu+1)\} +$$
$$\frac{i(n-1-k)(n-k)}{2n}c(n-1,k+1,\mu+1) -$$
$$\frac{i(n+k+2)(n+k+3)}{2(n+2)}[w+(2n+3)^2]c(n+1,k+1,\mu+1)$$
$$N_zc(n,k,\mu) = \frac{-i(k-\mu)(k+\mu)}{4(2k-1)(2k+1)}\{\frac{1}{n}c(n-1,k-1,\mu) -$$
$$\frac{1}{n+2}[w+(2n+3)^2]c(n+1,k-1,\mu)\} -$$
$$\frac{i(n-1-k)(n-k)}{n}c(n-1,k+1,\mu) +$$
$$\frac{i(n+k+2)(n+k+3)}{n+2}[w+(2n+3)^2]c(n+1,k+1,\mu) \qquad (4.13)$$

$$B_+c(n,k,\mu) = \frac{(k+\mu)(k+\mu-1)}{2(2k-1)(2k+1)}c(n,k-1,\mu-1) -$$
$$2(n-k)(n+2+k)c(n,k+1,\mu-1)$$
$$B_-c(n,k,\mu) = -\frac{(k-\mu)(k-\mu-1)}{2(2k-1)(2k+1)}c(n,k-1,\mu+1) +$$
$$2(n-k)(n+2+k)c(n,k+1,\mu+1)$$
$$B_zc(n,k,\mu) = -\frac{(k-\mu)(k+\mu)}{(2k-1)(2k+1)}c(n,k-1,\mu) -$$
$$4(n-k)(n+2+k)c(n,k+1,\mu)$$
$$J_+c(n,k,\mu) = (k+\mu)c(n,k,\mu-1) \quad J_-c(n,k,\mu) = (k-\mu)c(n,k,\mu+1)$$
$$J_zc(n,k,\mu) = 2\mu c(n,k,\mu) \qquad (4.14)$$

It is seen from the last expression that the meaning of the quantum number μ is such that $c(n,k,\mu)$ is the eigenfunction of the operator J_z with the eigenvalue 2μ, i.e. μ is the standard magnetic quantum number.

We use $\tilde{e}_{nk\mu}$ to denote basis vectors normalized to one and $\tilde{c}(n,k,\mu)$ to denote the wave function in the normalized basis. As follows from Eq. (4.12), the vectors $\tilde{e}_{nk\mu}$ can be defined as

$$\tilde{e}_{nk\mu} = \{(2k+1)!C_{2k}^{k-\mu}C_n^k C_{n+k+1}^k \prod_{j=1}^{n}[w+(2j+1)^2]\}^{-1/2} e_{nk\mu} \tag{4.15}$$

A direct calculation using Eqs. (4.12-4.15) shows that the action of the representation operators on the wave function in the normalized basis is given by

$$\mathcal{E}\tilde{c}(n,k,\mu) = \frac{1}{2}[\frac{(n-k)(n+k+1)}{n(n+1)}(w+(2n+1)^2)]^{1/2}\tilde{c}(n-1,k,\mu) +$$
$$\frac{1}{2}[\frac{(n+1-k)(n+k+2)}{(n+1)(n+2)}(w+(2n+3)^2)]^{1/2}\tilde{c}(n+1,k,\mu)]$$
$$N_+\tilde{c}(n,k,\mu) = \frac{i}{4}[\frac{(k+\mu)(k+\mu-1)}{(2k-1)(2k+1)(n+1)}]^{1/2}$$
$$\{[\frac{(n+k)(n+k+1)}{n}(w+(2n+1)^2)]^{1/2}\tilde{c}(n-1,k-1,\mu-1) -$$
$$[\frac{(n+2-k)(n+1-k)}{n+2}(w+(2n+3)^2)]^{1/2}\tilde{c}(n+1,k-1,\mu-1)\} -$$
$$\frac{i}{4}[\frac{(k+2-\mu)(k+1-\mu)}{(2k+1)(2k+3)(n+1)}]^{1/2}$$
$$\{[\frac{(n-k)(n-k-1)}{n}(w+(2n+1)^2)]^{1/2}\tilde{c}(n-1,k+1,\mu-1) -$$
$$[\frac{(n+k+2)(n+k+3)}{n+2}(w+(2n+3)^2)]^{1/2}\tilde{c}(n+1,k+1,\mu-1)]\}$$
$$N_-\tilde{c}(n,k,\mu) = -\frac{i}{4}[\frac{(k-\mu)(k-\mu-1)}{(2k-1)(2k+1)(n+1)}]^{1/2}$$
$$\{[\frac{(n+k)(n+k+1)}{n}(w+(2n+1)^2)]^{1/2}\tilde{c}(n-1,k-1,\mu+1) -$$
$$[\frac{(n+2-k)(n+1-k)}{n+2}(w+(2n+3)^2)]^{1/2}\tilde{c}(n+1,k-1,\mu+1)\} +$$
$$\frac{i}{4}[\frac{(k+2+\mu)(k+1+\mu)}{(2k+1)(2k+3)(n+1)}]^{1/2}$$
$$\{[\frac{(n-k)(n-k-1)}{n}(w+(2n+1)^2)]^{1/2}\tilde{c}(n-1,k+1,\mu+1) -$$
$$[\frac{(n+k+2)(n+k+3)}{n+2}(w+(2n+3)^2)]^{1/2}\tilde{c}(n+1,k+1,\mu+1)]\}$$
$$N_z\tilde{c}(n,k,\mu) = -\frac{i}{2}[\frac{(k-\mu)(k+\mu)}{(2k-1)(2k+1)(n+1)}]^{1/2}$$
$$\{[\frac{(n+k)(n+k+1)}{n}(w+(2n+1)^2)]^{1/2}\tilde{c}(n-1,k-1,\mu) -$$
$$[\frac{(n+2-k)(n+1-k)}{n+2}(w+(2n+3)^2)]^{1/2}\tilde{c}(n+1,k-1,\mu)\} -$$
$$\frac{i}{2}[\frac{(k+1-\mu)(k+1+\mu)}{(2k+1)(2k+3)(n+1)}]^{1/2}$$
$$\{[\frac{(n-k)(n-k-1)}{n}(w+(2n+1)^2)]^{1/2}\tilde{c}(n-1,k+1,\mu) -$$
$$[\frac{(n+k+2)(n+k+3)}{n+2}(w+(2n+3)^2)]^{1/2}\tilde{c}(n+1,k+1,\mu)]\} \tag{4.16}$$

$$B_+\tilde{c}(n,k,\mu) = [\frac{(k+\mu)(k+\mu-1)(n+1-k)(n+1+k)}{(2k-1)(2k+1)}]^{1/2}\tilde{c}(n,k-1,\mu-1)$$
$$-[\frac{(k+2-\mu)(k+1-\mu)(n-k)(n+k+2)}{(2k+1)(2k+3)}]^{1/2}\tilde{c}(n,k+1,\mu-1)$$
$$B_-\tilde{c}(n,k,\mu) = -[\frac{(k-\mu)(k-\mu-1)(n+1-k)(n+1+k)}{(2k-1)(2k+1)}]^{1/2}\tilde{c}(n,k-1,\mu+1)$$
$$+[\frac{(k+2+\mu)(k+1+\mu)(n-k)(n+k+2)}{(2k+1)(2k+3)}]^{1/2}\tilde{c}(n,k+1,\mu+1)$$
$$B_z\tilde{c}(n,k,\mu) = -2[\frac{(k-\mu)(k+\mu)(n+1-k)(n+1+k)}{(2k-1)(2k+1)}]^{1/2}\tilde{c}(n,k-1,\mu)$$
$$-2[\frac{(k+1-\mu)(k+1+\mu)(n-k)(n+k+2)}{(2k+1)(2k+3)}]^{1/2}\tilde{c}(n,k+1,\mu)$$
$$J_+\tilde{c}(n,k,\mu) = [(k+\mu)(k+1-\mu)]^{1/2}\tilde{c}(n,k,\mu-1)$$
$$J_-\tilde{c}(n,k,\mu) = [(k-\mu)(k+1+\mu)]^{1/2}\tilde{c}(n,k,\mu+1)$$
$$J_z\tilde{c}(n,k,\mu) = 2\mu\tilde{c}(n,k,\mu) \tag{4.17}$$

4.2 Semiclassical approximation

Consider now the semiclassical approximation in the $\tilde{e}_{nkl}$ basis. As noted in Secs. 3.2 and 3.6, the operator **B** is the dS analog of the usual momentum **P** such that in Poincare limit $\mathbf{B} = 2R\mathbf{P}$. The operator **J** has the same meaning as in Poincare invariant theory. Then it is clear from Eqs. (4.13) and (4.14) that a necessary condition for the semiclassical approximation is that the quantum numbers $(nk\mu)$ are much greater than 1 (in agreement with the remarks in the preceding section). By analogy with the discussion of the semiclassical approximation in Secs. 2.2 and 3.6, we assume that a state is semiclassical if its wave function has the form

$$\tilde{c}(n,k,\mu) = a(n,k,\mu)exp[i(-n\varphi + k\alpha - \mu\beta)] \tag{4.18}$$

where $a(n,k,\mu)$ is an amplitude, which is not small only in some vicinities of $n = n_0$, $k = k_0$ and $\mu = \mu_0$. We also assume that when the quantum numbers $(nk\mu)$ change by one, the main contribution comes from the rapidly oscillating exponent. Then, as follows from the first expression in Eq. (4.16), the action of the dS energy operator can be written as

$$\mathcal{E}\tilde{c}(n,k,\mu) \approx \frac{1}{n_0}[(n_0-k_0)(n_0+k_0)(w+4n_0^2)]^{1/2}cos(\varphi)\tilde{c}(n,k,\mu) \tag{4.19}$$

Therefore the semiclassical wave function is approximately the eigenfunction of the dS energy operator with the eigenvalue

$$\frac{1}{n_0}[(n_0-k_0)(n_0+k_0)(w+4n_0^2)]^{1/2}cos\varphi.$$

We will use the following notations. When we consider not the action of an operator on the wave function but its approximate eigenvalue in the semiclassical state, we will use for the eigenvalue the same notation as for the operator and this should not lead to misunderstanding. Analogously, in eigenvalues we will write n, k and μ instead of n_0, k_0 and μ_0, respectively. By analogy with Eq. (4.19) we can consider eigenvalues of the other operators and the results can be represented as

$$
\begin{aligned}
&\mathcal{E} = \frac{1}{n}[(n-k)(n+k)(w+4n^2)]^{1/2}cos\varphi \\
&N_x = (w+4n^2)^{1/2}\{-\frac{sin\varphi}{k}[\mu cos\alpha cos\beta + ksin\alpha sin\beta] + \\
&\frac{cos\varphi}{n}[\mu sin\alpha cos\beta - kcos\alpha sin\beta]\} \\
&N_y = (w+4n^2)^{1/2}\{-\frac{sin\varphi}{k}[\mu cos\alpha sin\beta - ksin\alpha cos\beta] + \\
&\frac{cos\varphi}{n}[\mu sin\alpha sin\beta + kcos\alpha cos\beta]\} \\
&N_z = [(k-\mu)(k+\mu)(w+4n^2)]^{1/2}(\frac{1}{k}sin\varphi cos\alpha - \frac{1}{n}cos\varphi sin\alpha) \\
&B_x = \frac{2}{k}[(n-k)(n+k)]^{1/2}[\mu cos\alpha cos\beta + ksin\alpha sin\beta] \\
&B_y = \frac{2}{k}[(n-k)(n+k)]^{1/2}[\mu cos\alpha sin\beta - ksin\alpha cos\beta] \\
&B_z = -\frac{2}{k}[(k-\mu)(k+\mu)(n-k)(n+k)]^{1/2}cos\alpha \\
&J_x = 2[(k-\mu)(k+\mu)]^{1/2}cos\beta \quad J_y = 2[(k-\mu)(k+\mu)]^{1/2}sin\beta \\
&J_z = 2\mu
\end{aligned} \tag{4.20}
$$

Since $\mathbf{B}$ is the dS analog of $\mathbf{p}$ and in classical theory $\mathbf{J} = \mathbf{r}\times\mathbf{p}$, one might expect that $\mathbf{BJ} = 0$ and, as follows from the above expressions, this is the case. It also follows that $\mathbf{B}^2 = 4(n^2-k^2)$ and $\mathbf{J}^2 = 4k^2$ in agreement with Eq. (4.11).

In Sec. 3.6 we described semiclassical wave functions by six parameters $(\mathbf{r},\mathbf{p})$ while in the basis $\tilde{e}_{nkl}$ the six parameters are $(n,k,\mu,\varphi,\alpha,\beta)$. Since in the dS theory the ten representation operators are on equal footing, it is also possible to describe a semiclassical state by semiclasscal eigenvalues of these operators. However, we should have four constraints for them. As follows from Eqs. (3.18) and (3.23), the constraints can be written as

$$\mathcal{E}^2 + \mathbf{N}^2 - \mathbf{B}^2 - \mathbf{J}^2 = w \quad \mathbf{N}\times\mathbf{B} = -\mathcal{E}\mathbf{J} \tag{4.21}$$

As noted in Sec. 3.6, in Poincare limit $\mathcal{E} = 2RE$, $\mathbf{B} = 2R\mathbf{p}$ (since we have replaced Eq. (1.4) by Eq. (4.1)) and the values of $\mathbf{N}$ and $\mathbf{J}$ are much less than $\mathcal{E}$ and $\mathbf{B}$. Therefore the first relation in Eq. (4.21) is the Poincare analog of the well-known relation $E^2 - \mathbf{p}^2 = m^2$.

The quantities $(nk\mu\varphi\alpha\beta)$ can be expressed in terms of semiclassical eigenvalues $(\mathcal{E}, \mathbf{N}, \mathbf{B}, \mathbf{J})$ as follows. The quantities $(nk\mu)$ can be found from the relations

$$\mathbf{B}^2 + \mathbf{J}^2 = 4n^2 \quad \mathbf{J}^2 = 4k^2 \quad J_z = 2\mu \tag{4.22}$$

and then the angles $(\varphi\alpha\beta)$ can be found from the relations

$$\begin{aligned}
&cos\varphi = \frac{2\mathcal{E}n}{B(w+4n^2)^{1/2}} \quad sin\varphi = -\frac{\mathbf{BN}}{B(w+4n^2)^{1/2}} \\
&cos\alpha = -JB_z/(BJ_\perp) \quad sin\alpha = (\mathbf{B}\times\mathbf{J})_z/(BJ_\perp) \\
&cos\beta = J_x/J_\perp \quad sin\beta = J_y/J_\perp
\end{aligned} \tag{4.23}$$

where $B = |\mathbf{B}|$, $J = |\mathbf{J}|$ and $J_\perp = (J_x^2 + J_y^2)^{1/2}$. In semiclassical approximation, uncertainties of the quantities $(nk\mu)$ should be such that $\Delta n \ll n$, $\Delta k \ll k$ and $\Delta\mu \ll \mu$. On the other hand, those uncertainties cannot be very small since the distribution in $(nk\mu)$ should be such that all the ten approximate eigenvalues $(\mathcal{E}, \mathbf{N}, \mathbf{B}, \mathbf{J})$ should be much greater than their corresponding uncertainties. The assumption is that for macroscopic bodies all these conditions can be satisfied.

In Sec. 3.6 we discussed operators in Poincare limit and corrections to them, which lead to the dS antigravity. A problem arises how the Poincare limit should be defined in the basis defined in the present chapter. In contrast to Sec. 3.6, we can now work not with the unphysical quantities $\mathbf{v}$ or $\mathbf{p} = m\mathbf{v}$ defined on the Lorentz hyperboloid but directly with semiclassical eigenvalues of the representation operators. In contrast to Sec. 3.6, we now *define* $\mathbf{p} = \mathbf{B}/(2R)$, $m = w^{1/2}/(2R)$ and $E = (m^2 + \mathbf{p}^2)^{1/2}$. Then Poincare limit can be defined by the requirement that when R is large, the quantities $\mathcal{E}$ and $\mathbf{B}$ are proportional to R while $\mathbf{N}$ and $\mathbf{J}$ do not depend on R. In this case, as follows from Eq. (4.21), in Poincare limit $\mathcal{E} = 2RE$ and $\mathbf{B} = 2R\mathbf{p}$.

4.3 Position operator in dS theory

By analogy with constructing a physical position operator in Sec. 2.11, the position operator in dS theory can be found from the following considerations. Since the operators $\mathbf{B}$ and $\mathbf{J}$ are consistently defined as representation operators of the dS algebra and we have defined $\mathbf{p}$ as $\mathbf{B}/2R$, one might seek the position operator such that on classical level the relation $\mathbf{r}\times\mathbf{p} = \mathbf{J}/2$ will take place (the factor 1/2 is a consequence of the fact that we work with units where $\hbar/2 = 1$). On classical level one can define parallel and perpendicular components of $\mathbf{r}$ as $\mathbf{r} = r_{||}\mathbf{B}/|\mathbf{B}| + \mathbf{r}_\perp$ and analogously $\mathbf{N} = N_{||}\mathbf{B}/|\mathbf{B}| + \mathbf{N}_\perp$. Then the relation $\mathbf{r}\times\mathbf{p} = \mathbf{J}/2$ defines uniquely only $\mathbf{r}_\perp$ and it follows from the second relation in Eq. (4.21) that $\mathbf{N}_\perp = -2E\mathbf{r}_\perp$. However, it is not clear yet how $r_{||}$ should be defined and whether the last relation is also valid for the parallel components of $\mathbf{N}$ and $\mathbf{r}$. As follows from the second relation in Eq. (4.23), it will be valid if $|sin\varphi| = r_{||}/R$, i.e. φ is the angular coordinate. As noted in Sec. 2.2, semiclassical approximation for a physical quantity can be valid only in states where this quantity is rather large. Therefore if R is very large then φ is very small if the distances are not cosmological (i.e. they are much less than R). Hence the problem arises whether this approximation is valid. This is a very important problem since in standard approach it is assumed that nevertheless φ can be considered semiclassically. Suppose first that this is the case and consider corrections to Poincare limit in classical limit.

Since $\mathbf{B} = 2R\mathbf{p}$ and $\mathbf{J}/2 = \mathbf{r}_\perp \times \mathbf{p}$ then it follows from Eq. (4.22) that in first order in $1/R^2$ we have $k^2/n^2 = \mathbf{r}_\perp^2/R^2$. Therefore as follows from the first expression in Eq. (4.20), in first order in $1/R^2$ the results on $\mathcal{E}$ and $\mathbf{N}$ can be represented as

$$\mathcal{E} = 2ER(1 - \frac{\mathbf{r}^2}{2R^2}), \quad \mathbf{N} = -2E\mathbf{r} \tag{4.24}$$

Hence the result for the energy is in agreement with Eq. (3.64) while the result for $\mathbf{N}$ is in agreement with Eq. (3.16).

Consider now constructing the position operator on quantum level. In view of the remarks in Sec. 4.1, we assume the approximation $n, k, |\mu| \gg 1$. Let us define Hermitian operators A and B which act as

$$\begin{aligned}
\mathcal{A}\tilde{c}(n,k,\mu) &= \frac{i}{2}[\tilde{c}(n+1,k,\mu) - \tilde{c}(n-1,k,\mu)] \\
\mathcal{B}\tilde{c}(n,k,\mu) &= \frac{1}{2}[\tilde{c}(n+1,k,\mu) + \tilde{c}(n-1,k,\mu)]
\end{aligned} \tag{4.25}$$

and the operators $\mathbf{F}$ and $\mathbf{G}$ which act as (compare with Eqs. (2.63) and (2.64))

$$\begin{aligned}
F_+\tilde{c}(n,k,\mu) &= -\frac{i}{4}[(k+\mu)\tilde{c}(n,k-1,\mu-1) + (k-\mu)\tilde{c}(n,k+1,\mu-1)] \\
F_-\tilde{c}(n,k,\mu) &= \frac{i}{4}[(k-\mu)\tilde{c}(n,k-1,\mu+1) + (k+\mu)\tilde{c}(n,k+1,\mu+1)] \\
F_z\tilde{c}(n,k,\mu) &= \frac{i}{2}\sqrt{k^2-\mu^2}[\tilde{c}(n,k-1,\mu) - \tilde{c}(n,k+1,\mu)]
\end{aligned} \tag{4.26}$$

$$\begin{aligned}
G_+\tilde{c}(n,k,\mu) &= \frac{1}{4k}[(k+\mu)\tilde{c}(n,k-1,\mu-1) - (k-\mu)\tilde{c}(n,k+1,\mu-1)] \\
G_-\tilde{c}(n,k,\mu) &= -\frac{1}{4k}[(k-\mu)\tilde{c}(n,k-1,\mu+1) - (k+\mu)\tilde{c}(n,k+1,\mu+1)] \\
G_z\tilde{c}(n,k,\mu) &= -\frac{\sqrt{k^2-\mu^2}}{2k}[\tilde{c}(n,k-1,\mu) + \tilde{c}(n,k+1,\mu)]
\end{aligned} \tag{4.27}$$

Then, as follows from Eqs. (4.16) and (4.17), the representation operators can be written as

$$
\begin{aligned}
&\mathcal{E}\tilde{c}(n,k,\mu) = \frac{\sqrt{n^2-k^2}}{n}(w+4n^2)^{1/2}B, \quad \mathbf{N} = -(w+4n^2)^{1/2}(\mathcal{A}\mathbf{G}+\frac{1}{n}\mathcal{B}\mathbf{F})\\
&\mathbf{B} = 2\sqrt{n^2-k^2}\mathbf{G}, \quad J_{\pm}\tilde{c}(n,k,\mu) = \sqrt{k^2-\mu^2}\tilde{c}(n,k,\mu\mp 1)\\
&J_z\tilde{c}(n,k,\mu) = 2\mu\tilde{c}(n,k,\mu) \qquad (4.28)
\end{aligned}
$$

and, as follows from Eqs. (4.26,7.15,4.28)

$$
\begin{aligned}
&[J_j,F_k] = 2ie_{jkl}F_l, \quad [J_j,G_k] = 2ie_{jkl}G_l, \quad \mathbf{G}^2 = 1, \quad \mathbf{F}^2 = k^2\\
&[G_j,G_k] = 0, \quad [F_j,F_k] = -\frac{i}{2}e_{jkl}J_l, \quad e_{jkl}\{F_k,G_l\} = J_j\\
&\mathbf{JG} = \mathbf{GJ} = \mathbf{JF} = \mathbf{FJ} = 0, \quad \mathbf{FG} = -\mathbf{GF} = i \qquad (4.29)
\end{aligned}
$$

The first two relations show that $\mathbf{F}$ and $\mathbf{G}$ are the vector operators as expected. The third relation shows that $\mathbf{G}$ can be treated as an operator of the unit vector along the direction of the momentum. The result for the anticommutator shows that on classical level $\mathbf{F}\times\mathbf{G} = \mathbf{J}/2$ and the last two relations show that on classical level the operators in the triplet $(\mathbf{F},\mathbf{G},\mathbf{J})$ are mutually orthogonal. Hence we have a full analogy with the corresponding results in Poincare invariant theory (see Sec. 2.11).

Let us define the operators $R_{||}$ and $\mathcal{R}_{\perp}$ as

$$
\mathcal{R}_{||} = R\mathcal{A}, \quad \mathcal{R}_{\perp} = \frac{R}{n}\mathbf{F} \qquad (4.30)
$$

Then taking into account that $(w+4n^2)^{1/2} = 2RE$, the expression for $\mathbf{N}$ in Eq. (4.28) can be written as

$$
\mathbf{N} = -2E\mathcal{R}_{||}\mathbf{G} - 2E\mathcal{B}\mathcal{R}_{\perp} \qquad (4.31)
$$

If the function $\tilde{c}(n,k,\mu)$ depends on φ as in Eq. (4.18) and φ is of the order of r/R then, as follows from Eq. (4.25), in the approximation when the terms of the order of $(r/R)^2$ in $\mathbf{N}$ can be neglected, $\mathcal{B}\approx 1$. In the approximation when n can be replaced by a continuous variable Rp

$$
\mathcal{R}_{||} = i\hbar\frac{\partial}{\partial p}, \quad \mathcal{R}_{\perp} = \frac{\hbar}{p}\mathbf{F} \qquad (4.32)
$$

where the dependence on $\hbar$ is restored. Hence in this approximation

$$
\mathbf{N} = -2E\mathcal{R}_{||}\mathbf{G} - 2E\mathcal{R}_{\perp} \qquad (4.33)
$$

and this result can be treated as an implementation of the decomposition $\mathbf{N} = N_{||}\mathbf{B}/|\mathbf{B}| + \mathbf{N}_{\perp}$ on the operator level. The semiclassical result $\mathbf{N} = -2E\mathbf{r}$ will take place if in semiclassical approximation $\mathcal{R}_{||}$ can be replaced by $r_{||}$ and $\mathcal{R}_{\perp}$ can be replaced by $\mathbf{r}_{\perp}$.

In the approximation when n can be replaced by the continuous variable Rp, the commutation relations between $\mathcal{R}_{||}$, different components of $\mathcal{R}_\perp$ and different components of $\mathbf{p} = p\mathbf{G}$ are the same as in Sec. 2.11. Hence the operators $\mathcal{R}_{||}$ and $\mathcal{R}_\perp$ can be treated as the parallel and transverse components of the position operator in dS theory. In particular, by analogy with the consideration in Chap. 2 we can conclude that in dS theory there is no WPS in directions transverse to $\mathbf{B}$ and there is no wave function in coordinate representation.

We now investigate the properties of the operators $\mathcal{A}$ and $\mathcal{B}$ since, as shown in the next chapter, these operators are present in the two-body mass and distance operators. The relations between the operators $\mathcal{A}$, $\mathcal{B}$ and n are

$$[\mathcal{A}, n] = i\mathcal{B} \quad [\mathcal{B}, n] = -i\mathcal{A} \quad [\mathcal{A}, \mathcal{B}] = 0 \quad \mathcal{A}^2 + \mathcal{B}^2 = 1 \tag{4.34}$$

As noted in Sec. 2.2, in standard quantum theory the semiclassical wave function in momentum space contains a factor $exp(-ipx)$. Since n is now the dS analog of pR, we assume that $\tilde{c}(n, k, \mu)$ contains a factor $exp(-in\varphi)$, i.e. the angle φ is the dS analof of $r_{||}/R$. It is reasonable to expect that since all the ten representation operators of the dS algebra are angular momenta, in dS theory one should deal only with angular coordinates which are dimensionless. If we assume that in semiclassical approximation the main contributions in $\mathcal{A}\tilde{c}(n, k, \mu)$ and $\mathcal{B}\tilde{c}(n, k, \mu)$ come from the rapidly oscillating exponent then

$$\mathcal{A}\tilde{c}(n, k, \mu) \approx sin\varphi\tilde{c}(n, k, \mu) \quad \mathcal{B}\tilde{c}(n, k, \mu) \approx cos\varphi\tilde{c}(n, k, \mu) \tag{4.35}$$

in agreement with the first two expressions in Eq. (4.23). Therefore φ is indeed the dS analog of $r_{||}/R$ and if $r_{||} \ll R$ we recover the result that $N_{||} \approx -2Er_{||}$. Eq. (4.35) can be treated in such a way that $\mathcal{A}$ is the operator of the quantity $sin\varphi$ and $\mathcal{B}$ is the operator of the quantity $cos\varphi$. However, the following question arises. As noted in Sect. 2.2, semiclassical approximation for a quantity can be correct only if this quantity is rather large. At the same time, we assume that $\mathcal{A}$ is the operator of the quantity which is very small if R is large.

If φ is small, we have $sin\varphi \approx \varphi$ and in this approximation $\mathcal{A}$ can be treated as the operator of the angular variable φ. This seems natural since, as shown in Sec. 2.11, in Poincare invariant theory the operator of the longitudinal coordinate is id/dp and $\mathcal{A}$ is the finite difference analog of derivative over n. When φ is not small, the argument that $\mathcal{A}$ is the operator of the quantity $sin\varphi$ follows. Since

$$arcsin\varphi = \sum_{l=0}^{\infty} \frac{(2l)!\varphi^{2l+1}}{4^l(l!)^2(2l+1)},$$

$$\Phi = \sum_{l=0}^{\infty} \frac{(2l)!\mathcal{A}^{2l+1}}{4^l(l!)^2(2l+1)}$$

can be treated as the operator of the quantity φ. Indeed, as follows from this expression and Eq. (4.34), $[\Phi, n] = i$ what is the dS analog of the relation $[\mathcal{R}_{||}, p] = i\hbar$ (see Sec. 2.11).

Chapter 5

Two-body systems in discrete basis

5.1 Two-body mass operator and the cosmological acceleration in discrete basis

Consider now a system of two free particles in dS theory. As follows from Eq. (3.18), in this case the Casimir operator of the second order is

$$I_2 = -\frac{1}{2}\sum_{ab}(M_{ab}^{(1)} + M_{ab}^{(2)})(M^{ab(1)} + M^{ab(2)}) \tag{5.1}$$

As explained in the preceding chapter, for our purposes spins of the particles can be neglected. Then, as follows from Eq. (4.5)

$$I_2 = w_1 + w_2 + 2\mathcal{E}_1\mathcal{E}_2 + 2\mathbf{N}_1\mathbf{N}_2 - 2\mathbf{B}_1\mathbf{B}_2 - 2\mathbf{J}_1\mathbf{J}_2 + 18 \tag{5.2}$$

where the subscripts 1 and 2 are used to denote operators for particle 1 and 2, respectively. By analogy with Eq. (4.5), one can define the two-body operator W, which is an analog of the quantity w:

$$I_2 = W - \mathbf{S}^2 + 9 \tag{5.3}$$

where $\mathbf{S}$ is the two-body spin operator which is the total angular momentum in the rest frame of the two-body system. Then, as follows from Eqs. (5.2) and (5.3),

$$W = w_1 + w_2 + 2(w_1 + 4n_1^2)^{1/2}(w_2 + 4n_2^2)^{1/2} - 2F - 2\mathbf{B}_1\mathbf{B}_2 - 2\mathbf{J}_1\mathbf{J}_2 + \mathbf{S}^2 + 9 \tag{5.4}$$

where in this chapter we use F to denote the operator

$$F = (w_1 + 4n_1^2)^{1/2}(w_2 + 4n_2^2)^{1/2} - \mathcal{E}_1\mathcal{E}_2 + 2\mathbf{N}_1\mathbf{N}_2 \tag{5.5}$$

Let I_{2P} be the Casimir operator of the second order in Poincare invariant theory. If E is the two-body energy operator in Poincare invariant theory and $\mathbf{P}$ is the two-body Poincare momentum then $I_{2P}=E^2-\mathbf{P}^2$. This operator is sometimes called the mass operator squared although in general I_{2P} is not positive definite (e.g. for tachyons). However, for macroscopic bodies it is positive definite, i.e. can be represented as M_0^2, the classical value of which is $M_0^2=m_1^2+m_2^2+2E_1E_2-2\mathbf{p}_1\mathbf{p}_2$. As follows from Eq. (5.4)

$$W=W_0-2F-2\mathbf{J}_1\mathbf{J}_2+\mathbf{S}^2+9 \tag{5.6}$$

where

$$W_0=w_1+w_2+2(w_1+4n_1^2)^{1/2}(w_2+4n_2^2)^{1/2}-2\mathbf{B}_1\mathbf{B}_2=4R^2M_0^2 \tag{5.7}$$

Consider first the case when semiclassical approximation is valid. In Sec. 3.6 we discussed operators in Poincare limit and corrections to them, which lead to the dS antigravity. A problem arises how the dS antigravity can be recovered in the discrete basis defined in the preceding chapter. Let us assume that the longitudinal part of the position operator is such that Eq. (4.24) is valid. Then as follows from Eq. (4.24), $F=2E_1E_2\mathbf{r}^2$ where $\mathbf{r}=\mathbf{r}_1-\mathbf{r}_2$. Let $M^2=W/4R^2$ be the mass squared in Poincare invariant theory with dS corrections. In the nonrelativistic approximation the last three terms in the r.h.s. of Eq. (5.6) can be neglected. Then if $M=m_1+m_2+H_{nr}$ where H_{nr} is the nonrelativistic Hamiltonian in the c.m. frame, it follows from Eq. (5.6) and the expression for F that in first order in $1/R^2$

$$H(\mathbf{r},\mathbf{q})=\frac{\mathbf{q}^2}{2m_{12}}-\frac{m_{12}\mathbf{r}^2}{2R^2} \tag{5.8}$$

i.e. the same result as that given by Eq. (3.68). As a consequence, the result for the cosmological acceleration obtained in the discrete basis is the same as in the basis discussed in Chap. 3. Note that the correction to the Hamiltonian is always negative and proportional to m_{12} in the nonrelativistic approximation.

In deriving the result given by Eq. (5.8), as well as in deriving the result given by Eq. (3.68), the notions of dS space, metric and connection have not been used. This is an independent argument that the cosmological acceleration is simply a kinematical effect in dS theory and can be explained without dark energy, empty space-time and other artificial notions.

Consider now a general case, i.e. we will not assume that Eq. (4.24) is necessarily valid. Then, as follows from Eq. (4.28)

$$F=(w_1+4n_1^2)^{1/2}(w_2+4n_2^2)^{1/2}G$$
$$G=1-\{\frac{1}{n_1n_2}[\sqrt{(n_1^2-k_1^2)(n_2^2-k_2^2)}+\mathbf{F}_1\mathbf{F}_2]\mathcal{B}_1\mathcal{B}_2+\mathcal{A}_1\mathcal{A}_2\mathbf{G}_1\mathbf{G}_2+$$
$$\frac{1}{n_1}\mathbf{F}_1\mathbf{G}_2\mathcal{B}_1\mathcal{A}_2+\frac{1}{n_2}\mathbf{G}_1\mathbf{F}_2\mathcal{A}_1\mathcal{B}_2\} \tag{5.9}$$

where the single-particle operators $(\mathcal{A}_j,\mathcal{B}_j,\mathbf{F}_j,\mathbf{G}_j)$ $(j=1,2)$ are defined in Sec. 4.3.

5.2 Two-body relative distance operator

In Sec. 4.3 we discussed semiclassical approximation for the single-particle position operator in dS theory. In this section we investigate how the relative distance operator can be defined in this theory. As already noted, among the operators of the dS algebra there are no operators which can be identified with the distance operator but there are reasons to think that in semiclassical approximation the values of E and $\mathbf{N}$ are given by Eq. (4.24). From the point of view of our experience in Poincare invariant theory, the dependence of E on $\mathbf{r}$ might seem to be unphysical since the energy depends on the choice of the origin. However, only invariant quantities have a physical meaning; in particular the two-body mass can depend only on relative distances which do not depend on the choice of the origin.

In view of Eq. (4.24) one might think that the operator $\tilde{\mathbf{D}} = \mathcal{E}_2\mathbf{N}_1 - \mathcal{E}_1\mathbf{N}_2$ might be a good operator which in semiclassical approximation is proportional to $E_1E_2\mathbf{r}$ at least in main order in $1/R^2$. However, the operator $\mathbf{D}$ defining the relative distance should satisfy the following conditions. First of all, it should not depend on the motion of the two-body system as a whole; in particular it should commute with the operator which is treated as a total momentum in dS theory. As noted in Sec. 4.1, the single-particle operator $\mathbf{J}'$ is a better candidate for the total momentum operator than $\mathbf{B}$. Now we use $\mathbf{J}'$ to denote the total two-particle operator $\mathbf{J}'_1+\mathbf{J}'_2$. Analogously, we use $\mathbf{J}$" to denote the total two-particle operator $\mathbf{J}_1$" + $\mathbf{J}_2$". As noted in Sec. 4.1, $\mathbf{J}$" can be treated as the internal angular momentum operator. Therefore, since $\mathbf{D}$ should be a vector operator with respect to internal rotations, it should properly commute with $\mathbf{J}$". In summary, the operator $\mathbf{D}$ should satisfy the relations

$$[J'^j, D^k] = 0 \quad [J"^j, D^k] = 2ie_{jkl}D^l \tag{5.10}$$

By using Eqs. (4.2) and (4.6) one can explicitly verify that the operator

$$\mathbf{D} = \mathcal{E}_2\mathbf{N}_1 - \mathcal{E}_1\mathbf{N}_2 - \mathbf{N}_1 \times \mathbf{N}_2 \tag{5.11}$$

indeed satisfies Eq. (5.10). If Poincare approximation is satisfied with a high accuracy then obviously $\mathbf{D} \approx \tilde{\mathbf{D}}$.

In contrast to the situation in standard quantum mechanics, different components of $\mathbf{D}$ do not commute with each other and therefore are not simultaneously measurable. As shown in Chap. 2, if in Poincare invariant theory the position operator is defined in a consistent way, its different components also do not commute with each other (see Sec. 2.11). However, since $[\mathbf{D}^2, \mathbf{J}"] = 0$, by analogy with quantum mechanics one can choose $(\mathbf{D}^2, \mathbf{J}"^2, J"_z)$ as a set of diagonal operators. The result of explicit calculations is

$$\mathbf{D}^2 = (\mathcal{E}_1^2 + \mathbf{N}_1^2)(\mathcal{E}_2^2 + \mathbf{N}_2^2) - (\mathcal{E}_1\mathcal{E}_2 + \mathbf{N}_1\mathbf{N}_2)^2 - 4(\mathbf{J}_1\mathbf{B}_2 + \mathbf{J}_2\mathbf{B}_1) - 4\mathbf{J}_1\mathbf{J}_2 \tag{5.12}$$

It is obvious that in typical situations the last two terms in this expression are much less than the first two terms and for this reason we accept an approximation

$$\mathbf{D}^2 \approx (\mathcal{E}_1^2 + \mathbf{N}_1^2)(\mathcal{E}_2^2 + \mathbf{N}_2^2) - (\mathcal{E}_1\mathcal{E}_2 + \mathbf{N}_1\mathbf{N}_2)^2 \tag{5.13}$$

Then, as follows from Eqs. (4.28, (5.5) and (5.9), in the approximation when $n_1, n_2 \gg 1$

$$\mathbf{D}^2 \approx (w_1 + 4n_1^2)(w_2 + 4n_2^2)(2 - G)G \qquad (5.14)$$

Hence the knowledge of the operator G is needed for calculating both, the two-body mass and distance operators.

At this point no assumption that semiclassical approximation is valid has been made. If Eq. (4.24) is valid then, as follows from Eq. (5.13), in first order in $1/R^2$ $\mathbf{D}^2 = 16E_1^2E_2^2R^2r^2$ where $r = |\mathbf{r}|$. In particular, in the nonrelativistic approximation $\mathbf{D}^2 = 16m_1^2m_2^2R^2r^2$, i.e. $\mathbf{D}^2$ is proportional to r^2 what justifies treating $\mathbf{D}$ as a dS analog of the relative distance operator.

By analogy with standard theory, we can consider the two-body system in its c.m. frame. Since we choose $\mathbf{B}+\mathbf{J}$ as the dS analog of momentum, the c.m. frame can be defined by the condition $\mathbf{B}_1+\mathbf{J}_2+\mathbf{B}_2+\mathbf{J}_2 = 0$. Therefore, as follows from Eq. (4.22), $n_1 = n_2$. This is an analog of the condition that the magnitudes of particle momenta in the c.m. frame are the same. Another simplification can be achieved if the position of particle 2 is chosen as the origin. Then $\mathbf{J}_2 = 0$, $\mathbf{J}_1 = (\mathbf{r}_\perp \times \mathbf{B}_1)/2R$, $B_2 = 2n_2$. In quantum theory these relations can be only approximate if semiclassical approximation is valid. Then, as follows from Eqs. (4.29) and (4.30), the expression for G in Eq. (5.9) has a much simpler form:

$$G = 1 - \frac{\sqrt{n_1^2 - k_1^2}}{n_1}(\mathcal{B}_1\mathcal{B}_2 - \mathcal{A}_1\mathcal{A}_2) \qquad (5.15)$$

In the approximation when $\mathcal{B}_i$ can be replaced by $cos\varphi_i$ and $\mathcal{A}_i$ - by $sin\varphi_i$ ($i = 1, 2$), we can again recover the above result $\mathbf{D}^2 = 16E_1^2E_2^2R^2r^2$ if $|\varphi_1 + \varphi_2| = r_{||}/R$ since $|\varphi_i| = |r_{||i}|/R$, $k_1^2/n_1^2 = r_\perp^2/R^2$ and the particle momenta are approximately antiparallel.

We conclude that if standard semiclassical approximation is valid then dS corrections to the two-body mass operator are of the order of $(r/R)^2$. This result is in agreement with standard intuition that dS corrections can be important only at cosmological distances while in the Solar System these corrections are negligible. On the other hand, as it has been already noted, those conclusions are based on belief that the angular distance φ, which is of the order of r/R, can be considered semiclassically in spite of the fact that it is very small. In the next section we investigate whether this is the case. Since from now on we are interested only in distances which are much less than cosmological ones, we will investigate what happens if all corrections of the order of r/R and greater are neglected. In particular, we accept the approximation that $|\mathbf{B}_1| = 2n_1$, $|\mathbf{B}_2| = 2n_2$ and the c.m. frame is defined by the condition $\mathbf{B}_1 + \mathbf{B}_2 = 0$.

By analogy with standard theory, it is convenient to consider the two-body mass operator if individual particle momenta n_1 and n_2 are expressed in terms of the total and relative momenta N and n. In the c.m. frame we can assume that $\mathbf{B}_1$ is directed along the positive direction of the z axis and then $\mathbf{B}_2$ is directed along the negative direction of the z axis. Therefore the quantum number N characterizing the total dS momentum can be defined as $N = n_1 - n_2$. In nonrelativistic theory the relative momentum is defined as $\mathbf{q} = (m_2\mathbf{p}_1 - m_1\mathbf{p}_2)/(m_1 + m_2)$ and in relativistic theory as $\mathbf{q} = (E_2\mathbf{p}_1 - E_1\mathbf{p}_2)/(E_1 + E_2)$. Therefore, taking into account the fact that in the c.m. frame the particle momenta are directed in opposite directions, one might define n as $n = (m_2n_1 + m_1n_2)/(m_1 + m_2)$ or $n = (E_2n_1 + E_1n_2)/(E_1 + E_2)$. These definitions involve Poincare masses and energies. Another possibility is $n = (n_1 + n_2)/2$. In all these cases we have that $n \to (n+1)$ when $n_1 \to (n_1+1)$, $n_2 \to (n_2+1)$ and $n \to (n-1)$ when $n_1 \to (n_1-1)$, $n_2 \to (n_2-1)$. In what follows, only this feature is important.

Although so far we are working in standard dS quantum theory over complex numbers, we will argue in the next chapters that fundamental quantum theory should be finite. We will consider a version of quantum theory where complex numbers are replaced by a Galois field. Let $\psi_1(n_1)$ and $\psi_2(n_2)$ be the functions describing the dependence of single-particle wave functions on n. Then in our approach only those functions $\psi_1(n_1)$ and $\psi_2(n_2)$ are physical which have a finite carrier in n_1 and n_2, respectively. Therefore we assume that $\psi_1(n_1)$ can be different from zero only if $n_1 \in [n_{1min}, n_{1max}]$ and analogously for $\psi_2(n_2)$. If $n_{1max} = n_{1min} + \delta_1 - 1$ then a necessary condition that n_1 is semiclassical is $\delta_1 \ll n_1$. At the same time, since δ_1 is the dS analog of $\Delta p_1 R$ and R is very large, we expect that $\delta_1 \gg 1$. We use ν_1 to denote $n_1 - n_{1min}$. Then if $\psi_1(\nu_1) = a_1(\nu_1)exp(-i\varphi_1\nu_1)$, we can expect by analogy with the consideration in Sect. 2.2 that the state $\psi_1(\nu_1)$ will be semiclassical if $|\varphi_1\delta_1| \gg 1$ since in this case the exponent makes many oscillations on $[0, \delta_1]$. Even this condition indicates that φ_1 cannot be extremely small. Analogously we can consider the wave function of particle 2, define δ_2 as the width of its dS momentum distribution and $\nu_2 = n_2 - n_{2min}$. The range of possible values of N and n is shown

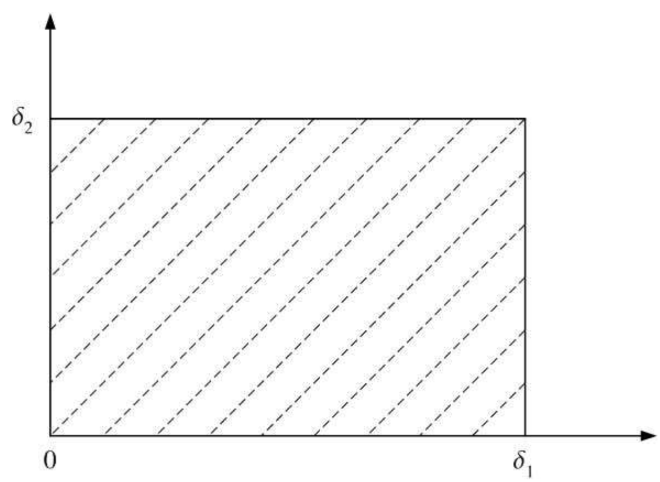

Figure 5.1: Range of possible values of N and n.

in Fig. 5.1 where it is assumed that $\delta_1 \geq \delta_2$. The minimum and maximum values of

N are $N_{min} = n_{1min} - n_{2max}$ and $N_{max} = n_{1max} - n_{2min}$, respectively. Therefore N can take $\delta_1 + \delta_2$ values. Each incident dashed line represents a set of states with the same value of N and different values of n. We now use n_{min} and n_{max} to define the minimum and maximum values of the relative dS momentum n. For each fixed value of N those values are different, i.e. they are functions of N. Let $\delta(N) = n_{max} - n_{min}$ for a given value of N. It is easy to see that $\delta(N) = 0$ when $N = N_{min}$ and $N = N_{max}$ while for other values of N, $\delta(N)$ is a natural number in the range $(0, \delta_{max}]$ where $\delta_{max} = min(\delta_1, \delta_2)$. The total number of values of (N, n) is obviously $\delta_1\delta_2$, i.e.

$$\sum_{N=Nmin}^{Nmax} \delta(N) = \delta_1\delta_2 \tag{5.16}$$

As follows from Eq. (4.25)

$$(\mathcal{B}_1\mathcal{B}_2 - \mathcal{A}_1\mathcal{A}_2)\psi_1(n_1)\psi_2(n_2) = \frac{1}{2}[\psi_1(n_1+1)\psi_2(n_2+1) + \psi_1(n_1-1)\psi_2(n_2-1)] \tag{5.17}$$

Therefore in terms of the variables N and n

$$(\mathcal{B}_1\mathcal{B}_2 - \mathcal{A}_1\mathcal{A}_2)\psi(N, n) = \frac{1}{2}[\psi(N, n+1) + \psi(N, n-1)] \tag{5.18}$$

Hence the operator $(\mathcal{B}_1\mathcal{B}_2 - \mathcal{A}_1\mathcal{A}_2)$ does not act on the variable N while its action on the variable n is described by the same expressions as the actions of the operators $\mathcal{B}_i$ $(i = 1, 2)$ on the corresponding wave functions. Therefore, considering the two-body system, we will use the notation $\mathcal{B} = \mathcal{B}_1\mathcal{B}_2 - \mathcal{A}_1\mathcal{A}_2$ and formally the action of this operator on the internal wave function is the same as in the second expression in Eq. (4.25). With this notation and with neglecting terms of the order of r/R and higher, Eqs. (5.6) and (5.15) can be written as

$$G = 1 - \mathcal{B}, \quad W = W_0 - 2(w_1 + 4n_1^2)^{1/2}(w_2 + 4n_2^2)^{1/2}G \tag{5.19}$$

Since both, the operator $\mathbf{D}^2$ and the dS correction to the operator W are defined by the same operator G, physical quantities corresponding to $\mathbf{D}^2$ and W will be semiclassical or not depending on whether the quantity corresponding to G is semiclassical or not. As follows from Eq. (4.34), the spectrum of the operator $\mathcal{B}$ can be only in the range [0,1] and therefore, as follows from Eq. (5.19), the same is true for the spectrum of the operator G. Hence, as follows from Eq. (5.19), any dS correction to the operator W is negative and in the nonrelativistic approximation is proportional to particle masses.

5.3 Validity of semiclassical approximation

Since classical mechanics works with a very high accuracy at macroscopic level, one might think that the validity of semiclassical approximation at this level is beyond any doubts. However, to the best of our knowledge, this question has not been investigated quantitavely. As discussed in Sect. 2.2, such quantities as coordinates and momenta are semiclassicall if their uncertainties are much less than the corresponding mean values. Consider wave functions describing the motion of macroscopic bodies as a whole (say the wave functions of the Sun, the Earth, the Moon etc.). It is obvious that uncertainties of coordinates in these wave functions are much less than the corresponding macroscopic dimensions. What are those uncertainties for the Sun, the Earth, the Moon, etc.? What are the uncertainties of their momenta? In standard quantum mechanics, the validity of semiclassical approximation is defined by the product $\Delta r \Delta p$ while each uncertainty by itself can be rather large. On the other hand, as shown in Chap. 2, the standard position operator should be reconsidered. Do we know what scenario for the distribution of momenta and coordinates takes place for macroscopic bodies?

In this section we consider several models of the function $\psi(n)$ where it is be possible to explicitly calculate $\bar{G}$ and ΔG and check whether the condition $\Delta G \ll |\bar{G}|$ (showing that the quantity G in the state ψ is semiclassical) is satisfied. As follows from Eq. (4.34), $[G, n] = i\mathcal{A}$ where formally the action of this operator on the internal wave function is the same as in the first expression in Eq. (4.25). Therefore, as follows from Eq. (2.2), $\Delta G \Delta n \geq \bar{\mathcal{A}}/2$.

As noted in Sect. 2.2, one might think that a necessary condition for the validity of semiclassical approximation is that the exponent in the semiclassical wave function makes many oscillations in the region where the wave function is not small. We will consider wave functions $\psi(n)$ containing $exp(-i\varphi n)$ such that $\psi(n)$ can be different from zero only if $n \in [n_{min}, n_{max}]$. Then, if $\delta = n_{max} - n_{min} + 1$, the exponent makes $|\varphi|\delta/2\pi$ oscillations on $[n_{min}, n_{max}]$ and φ should satisfy the condition $|\varphi| \gg 1/\delta$. The problem arises whether this condition is sufficient.

Our first example is such that $\psi(n) = exp(-i\varphi n)/\delta^{1/2}$ if $n \in [n_{min}, n_{max}]$. Then a simple calculation gives

$$\bar{G} = 1 - cos\varphi + \frac{1}{\delta}cos\varphi, \quad \Delta G = \frac{(\delta-1)^{1/2}cos\varphi}{\delta}, \quad \bar{\mathcal{A}} = (1 - \frac{1}{\delta})sin\varphi$$
$$\bar{n} = (n_{min} + n_{max})/2, \quad \Delta n = \delta(\frac{1 - 1/\delta^2}{12})^{1/2} \tag{5.20}$$

Since φ is of the order of r/R, we will always assume that $\varphi \ll 1$. Therefore for the validity of the condition $\Delta G \ll \bar{G}$, $|\varphi|$ should be not only much greater than $1/\delta$ but even much greater than $1/\delta^{1/4}$. Note also that $\Delta G \Delta n$ is of the order of $\delta^{1/2}$, i.e. much greater than $\bar{\mathcal{A}}$. This result shows that the state $\psi(\nu)$ is strongly non-semiclassical. The calculation shows that for ensuring the validity of semiclassical approximation, one should consider functions $\psi(\nu)$ which are small when n is close to n_{min} or n_{max}.

The second example is $\psi(\nu) = const\, C_\delta^\nu exp(-i\varphi\nu)$ where $\nu = n - n_{min}$ and *const* can be defined from the normalization condition. Since $C_\delta^\nu = 0$ when $\nu < 0$ or $\nu > \delta$, this function is not zero only when $\nu \in [0, \delta]$. The result of calculations is that $const^2 = 1/C_{2\delta}^{\delta}$ and

$$\bar{G} = 1 - cos\varphi + \frac{cos\varphi}{\delta+1}, \quad \Delta G = [\frac{sin^2\varphi}{\delta+1} + \frac{2}{\delta^2} + O(\frac{1}{\delta^3})]^{1/2}, \quad \bar{\mathcal{A}} = \frac{\delta sin\varphi}{\delta+1}$$
$$\bar{n} = \frac{1}{2}(n_{min} + n_{max}), \quad \Delta n = \frac{\delta}{2(2\delta-1)^{1/2}} \quad (5.21)$$

Now for the validity of the condition $\Delta G \ll \bar{G}$, $|\varphi|$ should be much greater than $1/\delta^{1/2}$ and $\Delta G \Delta n$ is of the order of $|\bar{\mathcal{A}}|$ which shows that the function is semiclassical. The matter is that $\psi(\nu)$ has a sharp peak at $\nu = \delta/2$ and by using Stirling's formula it is easy to see that the width of the peak is of the order of $\delta^{1/2}$. It is also clear from the expression for $\bar{G}$ that this quantity equals the semiclassical value $1 - cos\varphi$ with a high accuracy only when $|\varphi| \gg 1/\delta^{1/2}$. This example might be considered as an indication that a semiclassical wave function such that the condition $|\varphi| \gg 1/\delta$ is sufficient, should satisfy the following properties. On one hand the width of the maximum should be of the order of δ and on the other the function should be small when n is close to n_{min} or n_{max}.

In view of this remark, the third example is $\psi(\nu) = const\, exp(-i\varphi\nu)\nu(\delta - \nu)$ if $n \in [n_{min}, n_{max}]$. Then the normalization condition is $const^2 = [\delta(\delta^4 - 1)/30]^{-1}$ and the result of calculations is

$$\bar{G} = 1 - cos\varphi + \frac{5cos\varphi}{\delta^2} + O(\frac{1}{\delta^3}), \quad \bar{\mathcal{A}} = sin\varphi\,(1 - \frac{5}{\delta^2}), \quad \bar{n} = (n_{min} + n_{max})/2$$
$$\overline{G^2} = (1 - cos\varphi)^2 + \frac{10}{\delta^2}(cos\varphi - cos2\varphi) + \frac{15cos\varphi}{\delta^3} + O(\frac{1}{\delta^4})$$
$$\Delta G = \frac{1}{\delta}[10sin^2\varphi + \frac{15cos\varphi}{\delta} + O(\frac{1}{\delta^2})]^{1/2}, \quad \Delta n = \frac{\delta}{2\sqrt{7}} \quad (5.22)$$

Now $\bar{G} \approx 1 - cos\varphi$ if $|\varphi| \gg 1/\delta$ but $\Delta G \ll |\bar{G}|$ only if $|\varphi| \gg 1/\delta^{3/4}$ and $\Delta G \Delta n$ is of the order of $|\bar{\mathcal{A}}|$ only if $|\varphi| \gg 1/\delta^{1/2}$. The reason why the condition $|\varphi| \gg 1/\delta$ is not sufficient is that $\overline{G^2}$ approximately equals its classical value $(1 - cos\varphi)^2$ only when $|\varphi| \gg 1/\delta^{3/4}$. The term with $1/\delta^3$ in $\overline{G^2}$ arises because when ν is close to 0, $\psi(\nu)$ is proportional only to the first degree of ν and when ν is close to δ, it is proportional to $\delta - \nu$.

Our last example is $\psi(\nu) = const\, exp(-i\varphi\nu)[\nu(\delta - \nu)]^2$ if $n \in [n_{min}, n_{max}]$. It will suffice to estimate sums $\sum_{\nu=1}^{\delta} \nu^k$ by $\delta^{k+1}/(k+1) + O(\delta^k)$. In particular, the normalization condition is $const^2 = 35 \cdot 18/\delta^9$ and the result of calculations is

$$\bar{G} = 1 - cos\varphi + \frac{6cos\varphi}{\delta^2} + O(\frac{1}{\delta^4}), \quad \bar{\mathcal{A}} = sin\varphi\,(1 - \frac{6}{\delta^2}), \quad \bar{n} = (n_{min} + n_{max})/2$$
$$\overline{G^2} = (1 - cos\varphi)^2 + \frac{12}{\delta^2}(cos\varphi - cos2\varphi) + O(\frac{1}{\delta^4})$$
$$\Delta G = \frac{1}{\delta}[12sin^2\varphi + O(\frac{1}{\delta^2})]^{1/2}, \quad \Delta n = \frac{\delta}{2\sqrt{11}} \quad (5.23)$$

In this example the condition $|\varphi| \gg 1/\delta$ is sufficient to ensure that $\Delta G \ll |\bar{G}|$ and $\Delta G \Delta n$ is of the order of $|\bar{\mathcal{A}}|$.

At the same time, the following question arises. If we wish to perform mathematical operations with a physical quantity in classical theory, we should guarantee that not only this quantity is semiclassical but a sufficient number of its powers is semiclassical too. Since the classical value of G is proportional to φ^2 and φ is small, there is no guaranty that for the quantity G this is the case. Consider, for example, whether G^2 is semiclassical. It is clear from Eq. (5.23) that $\overline{G^2}$ is close to its classical value $(1-cos\varphi)^2$ if $|\varphi| \gg 1/\delta$. However, $\Delta(G^2)$ will be semiclassical only if $\overline{G^4}$ is close to its classical value $(1-cos\varphi)^4$. A calculation with the wave function from the last example gives

$$\overline{G^4} = (1-cos\varphi)^4 + \frac{24}{\delta^2}(1-cos\varphi)^3(3+4cos\varphi) +$$
$$\frac{84}{\delta^4}(1-cos\varphi)^2(64cos^2\varphi + 11cos\varphi - 6) + \frac{35\cdot 9}{2\delta^5} + O(\frac{1}{\delta^6}) \qquad (5.24)$$

Therefore $\overline{G^4}$ will be close to its classical value $(1-cos\varphi)^4$ only if $|\varphi| \gg 1/\delta^{5/8}$. Analogously, if $\psi(\nu) = const[\nu(\delta-\nu)]^3$ then G^2 will be semiclassical but G^3 will not. This consideration shows that a sufficient number of powers of G will be semiclassical only if $\psi(n)$ is sufficiently small in vicinities of n_{min} and n_{max}. On the other hand, in the example described by Eq. (5.21), the width of maximum is much less than δ and therefore the condition $|\varphi| \gg 1/\delta$ is still insufficient.

The problem arises whether it is possible to find a wave function such that the contributions of the values of ν close to 0 or δ is negligible while the effective width of the maximum is or order δ. For example, it is known that for any segment $[a,b]$ and any $\epsilon < (b-a)/2$ it is possible to find an infinitely differentiable function $f(x)$ on $[a,b]$ such that $f(x)=0$ if $x \notin [a,b]$ and $f(x)=1$ if $x \in [a+\epsilon, b-\epsilon]$. However, we cannot use such functions for several reasons. First of all, the values of ν can be only integers: $\nu = 0,1,2,...\delta$. Another reason is that for correspondence with GFQT we can use only rational functions and even $exp(-i\nu\varphi)$ should be expressed in terms of rational functions (see Sec. 6.1).

In view of this discussion, we accept that the functions similar to that described in the second example give the best approximation for semiclassical approximation since in that case it is possible to prove that the condition $|\varphi| \gg 1/\delta^{1/2}$ guarantees that sufficiently many quantities G^k $(k=1,2,...)$ will be semiclassical. The first step of the proof is to show by induction that

$$G^k\psi(\nu) = \frac{(-1)^k}{2^k}\sum_{l=0}^{2k} C_{2k}^l (-1)^l \psi(\nu+k-l) \qquad (5.25)$$

Then the calculation of the explicit expression for $\overline{G^k}$ involves hypergeometric functions

$$F(-\delta, -\delta + k; k + 1; 1) = \sum_{l=0}^{\infty} \frac{(-\delta)_l(-\delta + k)_l}{l!(k + 1)_l}$$

where $(k)_l$ is the Pochhammer symbol. Such sums are finite and can be calculated by using the Saalschutz theorem [91]: $F(-\delta, -\delta + k; k + 1; 1) = k!(2\delta + k)!/\delta!(\delta + k)!$. As a result,

$$\overline{G^k} = \frac{(-1)^k(\delta!)^2 exp(-i\varphi k)}{2^k(\delta + k)!(\delta - k)!} F(-2k, -\delta - k; \delta - k + 1; exp[i(\varphi + \pi)]) \tag{5.26}$$

The hypergeometric function in this expression can be rewritten by using the formula [91]

$$F(a, b; 1 + a - b; z) = (1 + z)^{-a} F[\frac{a}{2}, \frac{a + 1}{2}; 1 + a - b; \frac{4z}{(1 + z)^2}]$$

As a consequence

$$\overline{G^k} = \frac{2^k(\delta!)^2}{(\delta + k)!(\delta - k)!} \sum_{l=0}^{k} \frac{(-k)_l(-k + \frac{1}{2})_l}{l!(\delta + 1 - k)_l} (sin\frac{\varphi}{2})^{2(k-l)} \tag{5.27}$$

This result shows that $\overline{G^k}$ is given by a series in powers of $1/[\delta sin^2(\varphi/2)]$. Hence if $\varphi \ll 1$ but $|\varphi| \gg 1/\delta^{1/2}$ we get that the classical expression for $\overline{G^k}$ is $(\overline{G^k})_{class} = 2^k sin^{2k}(\varphi/2)$ and the semiclassical approximation for G^k is valid since if $k \ll \delta$ then

$$\frac{\Delta(G^k)}{\overline{G^k}} = \frac{(2k^2 - k)^{1/2}}{\delta^{1/2} sin(\varphi/2)} + O(\frac{1}{\delta sin^2(\varphi/2)}) \tag{5.28}$$

Since φ is of the order of r/R, the condition $|\varphi| \gg 1/\delta^{1/2}$ is definitely satisfied at cosmological distances while the problem arises whether it is satisfied in the Solar System. Since δ can be treated as $2R\Delta q$ where Δq is the width of the relative momentum distribution in the internal two-body wave function, $\varphi\delta$ is of the order of $r\Delta q$. For understanding what the order of magnitude of this quantity is, one should have estimations of Δq for macroscopic wave functions. However, to the best of our knowledge, the existing theory does not make it possible to give reliable estimations of this quantity.

Below we argue that Δq is of the order of $1/r_g$ where r_g is the gravitational (Schwarzschild) radius of the component of the two-body system which has the greater mass. Then $\varphi\delta$ is of the order of r/r_g. This is precisely the parameter defining when standard Newtonian gravity is a good approximation to GR. For example, the gravitational radius of the Earth is of the order of $0.01m$ while the radius of the Earth is $R_E = 6.4 \times 10^6 m$. Therefore R_E/r_g is of the order of 10^9. The gravitational radius of the Sun is of the order of $3000m$, the distance from the Sun to the Earth is or order $150 \times 10^9 m$ and so r/r_g is of the order of 10^8. At the same time, the above discussion shows that the condition $\varphi\delta \gg 1$ is not sufficient for ensuring semiclassical approximation while the condition $|\varphi| \gg 1/\delta^{1/2}$ is. Hence we should compare the quantities r/R and $(r_g/R)^{1/2}$. Then it is immediately clear that the requirement $|\varphi| \gg 1/\delta^{1/2}$ will not be satisfied if R is very large. For example, if R is of the order of $10^{26}m$ then in the example with the Earth r/R is of the order of 10^{-19} and $(r_g/R)^{1/2}$ is of the order of 10^{-14} while in the example with the Sun r/R is of the order of 10^{-15} and $(r_g/R)^{1/2}$ is of the order of 10^{-10}. Therefore in these examples the requirement $|\varphi| \gg 1/\delta^{1/2}$ is not satisfied.

Our concusion is as follows. As shown in Chap. 2, even in standard Poincare invariant theory the position operator should be defined not by the set $(i\hbar\partial/\partial p_x, i\hbar\partial/\partial p_y, i\hbar\partial/\partial p_z)$ but by the operators $(\mathcal{R}_{||}, \mathcal{R}_{\perp})$. At the same time, the distance operator can be still defined in the standard way, i.e. by the operator $-\hbar^2(\partial/\partial\mathbf{p})^2$. However, explicit examples discussed in this section show that for macroscopic bodies semiclassical approximation can be valid only if standard distance operator is modified too.

5.4 Distance operator for macroscopic bodies

As noted in Chap. 2, standard position operator in quantum theory is defined by the requirement that the momentum and coordinate representations are related to each other by a Fourier transform and this requirement is postulated by analogy with classical electrodynamics. However, as discussed in Chap. 2, the validity of such a requirement is problematic and there exist situations when standard position operator does not work. In addition, in Poincare invariant theories there is no parameter R; in particular rapidly oscillating exponents do not contain this parameter.

In the case of macroscopic bodies a new complication arises. It will be argued in the next chapters that in GFQT the width δ of the n-distribution for a macroscopic body is inversely proportional to its mass. Therefore for nuclei and elementary particles the quantity δ is much greater than for macroscopic bodies and the requirement $|\varphi| \gg 1/\delta^{1/2}$ can be satisfied in some situations. On the other hand, such a treatment of the distance operator for macroscopic bodies is incompatible with semiclassical approximation since, as discussed in the preceding section, if the distances are not cosmological then φ is typically much less than $1/\delta^{1/2}$. Hence the interpretation of the distance operator for macroscopic bodies should be modified.

As noted in Secs. 2.2 and 2.3, in standard theory the semiclassical wave function in momentum space has the form $exp(-i\mathbf{r}\mathbf{p})a(\mathbf{p})$ where the amplitude $a(\mathbf{p})$ has a sharp maximum at the classical value of momentum $\mathbf{p} = \mathbf{p}_0$ and $\mathbf{r}$ is the classical radius-vector. This property is based on the fact that in standard theory the coordinate and momentum representations are related to each other by the Fourier transform. However, as shown in Chap. 2, the standard position operator should be modified and hence the problem of the form of the semiclassical wave function should be reconsidered. In this section we discuss how the semiclassical wave function in the

n-representation should depend on the classical value φ.

As noted in Sec. 2.2, a necessary condition for semiclassical approximation is that the wave function should make many oscillations in the region where its amplitude is not negligible. Hence if the rapidly oscillating exponent in the wave function is $exp(-i\varphi n)$ then the number of oscillations is of the order of $\varphi\delta$ and this number is large if $\varphi \gg 1/\delta$. As noted in the preceding section, this condition is typically satisfied but for the validity of semiclassical approximation the value of φ should be not only much greater than $1/\delta$ but even much greater than $1/\delta^{1/2}$. We assume that in the general case rapidly oscillating exponent in the wave function is not $exp(-i\varphi n)$ but $exp(-i\chi n)$ where χ is a function of φ such that $\chi(\varphi) = \varphi$ when $\varphi \gg 1/\delta^{1/2}$ (in particular when φ is of the order of cosmological distances) while for macroscopic bodies in the Solar System (when φ is very small), χ is a function of $\varphi = r/R$ to be determined. Note that when we discussed the operator $\mathbf{D}^2$ compatible with the standard interpretation of the distance operator, we did not neglect $\mathbf{J}$ in this operator and treated $|\varphi|$ as $r_{||}/R$. However, when we neglect all corrections of the order of $1/R$ and higher, we neglect $\mathbf{J}$ in $\mathbf{D}^2$ and replace φ by χ which does not vanish when $R \to \infty$. As shown in Sect. 5.2, the operator $\mathbf{D}^2$ is rotationally invariant since the internal two-body momentum operator is a reduction of the operator $\mathbf{J}"$ on the two-body rest states, $\mathbf{D}$ satisfies Eq. (5.10) and therefore $[\mathbf{J}", \mathbf{D}^2] = 0$. Hence χ can be only a function of r but not $r_{||}$.

Ideally, a physical interpretation of an operator of a physical quantity should be obtained from the quantum theory of measurements which should describe the operator in terms of a measurement of the corresponding physical quantity. However, although quantum theory is known for 80+ years, the quantum theory of measurements has not been developed yet. Our judgment about operators of different physical quantities can be based only on intuition and comparison of theory and experiment. As noted in Sect. 2.2, in view of our macroscopic experience, it seems unreasonable that if the uncertainty Δr of r does not depend on r then the relative accuracy $\Delta r/r$ in the measurement of r is better when r is greater.

When $exp(-i\varphi n)$ is replaced by $exp(-i\chi n)$, the results obtained in the preceding section remain valid but φ should be replaced by χ. Suppose that when φ is of the order of $1/\delta^{1/2}$ or less, $\chi = f(C(\varphi\delta)^\alpha)$ where C is a constant and $f(x)$ is a function such that $f(x) = x + o(x)$ where the correction $o(x)$ will be discussed later. Then if χ and φ are treated not as classical but as quantum physical quantities we have that $\Delta\chi \approx C\varphi^{\alpha-1}\delta^\alpha\Delta\varphi$. If φ is replaced by χ then, as follows from the first expression in Eq. (5.21), if $\chi \gg 1/\delta^{1/2}$ and $\chi \ll 1$, the operator G can be treated as the operator of the quantity $\chi^2/2$. Then it follows from the second expression in Eq. (5.21) that $\Delta(\chi^2)$ is of the order of $\chi/\delta^{1/2}$ and therefore $\Delta\chi$ is of the order of $1/\delta^{1/2}$. As a consequence, $\Delta\varphi \approx const \cdot \varphi(\varphi\delta)^{-\alpha}/\delta^{1/2}$. Since ($\varphi \gg 1/\delta$), the accuracy of the measurement of φ is better when $\alpha < 0$. In that case the relative accuracy $\Delta\varphi/\varphi$ is better for lesser values of φ and, as noted in Sect. 2.2, this is a desired behavior in view of our macroscopic experience. Note also that the condition $\alpha < 0$ is natural from the fact that $\chi \ll 1$ is a necessary condition for the wave function in momentum representation to be approximately continuous since the standard momentum is of the order of n/R.

If $\alpha < 0$ then $\Delta\varphi \approx const \cdot \varphi(\varphi\delta)^{|\alpha|}/\delta^{1/2}$. In view of quantum mechanical experience, one might expect that the accuracy should be better if δ is greater. On the other hand, in our approach δ is inversely proportional to the masses of the bodies under consideration and our macroscopic experience tells us that the accuracy of the measurement of relative distance does not depend on the mass. Indeed, suppose that we measure a distance by sending a light signal. Then the accuracy of the measurement should not depend on whether the signal is reflected by the mass $1kg$ or $1000kg$. Therefore at macroscopic level the accuracy should not depend on δ. Hence the optimal choice is $\alpha = -1/2$. In that case $\Delta\varphi \approx const \cdot \varphi^{3/2}$ and $\chi = f(C/(\varphi\delta)^{1/2})$. Then, if C is of the order of unity, the condition $\chi \gg 1/\delta^{1/2}$, which, as explained in the preceding section, guarantees that semiclassical approximation is valid, is automatically satisfied since in the Solar System we always have $(R/r)^{1/2} \gg 1$. We will see in the next section that such a dependence of χ on φ and δ gives a natural explanation of the Newton law of gravity.

5.5 Newton's law of gravity

As follows from Eqs. (5.21), with φ replaced by χ, the mean value of the operator G is $1 - cos\chi$ with a high accuracy. Consider two-body wave functions having the form $\psi(N,n) = [\delta(N)/(\delta_1\delta_2)]^{1/2}\psi(n)$. As follows from Eq. (5.16), such functions are normalized to one. Then, as follows from Eq. (5.19), the mean value of the operator W can be written as

$$\overline{W} = 4R^2M_0^2 + \overline{\Delta W}, \quad \overline{\Delta W} = -2[(w_1+4n_1^2)(w_2+4n_2^2)]^{1/2}F(\delta_1,\delta_2,\varphi)$$
$$F(\delta_1,\delta_2,\varphi) = \frac{1}{\delta_1\delta_2}\sum_{N=Nmin}^{Nmax}\delta(N)\{1-cos[f(\frac{C}{(\varphi\delta(N))^{1/2}})]\} \tag{5.29}$$

Strictly speaking, the semiclassical form of the wave function $exp(-i\chi n)a(n)$ cannot be used if $\delta(N)$ is very small; in particular, it cannot be used when $\delta(N) = 0$. We assume that in these cases the internal wave function can be modified such that the main contribution to the sum in Eq. (5.29) is given by those N where $\delta(N)$ is not small.

If φ is so large that the argument α of cos in Eq. (5.29) is extremely small, then the correction to Poincare limit is negligible. The next approximation is that this argument is small such we can approximate $cos(\alpha)$ by $1 - \alpha^2/2$. Then, taking into account that $f(\alpha) = \alpha + o(\alpha)$ and that the number of values of N is $\delta_1 + \delta_2$ we get

$$\overline{\Delta W} = -C^2[(w_1+4n_1^2)(w_2+4n_2^2)]^{1/2}\frac{\delta_1+\delta_2}{\delta_1\delta_2|\varphi|} \tag{5.30}$$

Now, by analogy with the derivation of Eq. (5.8), it follows that the classical nonrelativistic Hamiltonian is

$$H(\mathbf{r},\mathbf{q}) = \frac{\mathbf{q}^2}{2m_{12}} - \frac{m_1 m_2 R C^2}{2(m_1+m_2)r}(\frac{1}{\delta_1}+\frac{1}{\delta_2}) \qquad (5.31)$$

We see that the correction disappears if the width of the dS momentum distribution for each body becomes very large. In standard theory (over complex numbers) there is no limitation on the width of distribution while, as noted in the preceding section, in semiclassical approximation the only limitation is that the width of the dS momentum distribution should be much less than the mean value of this momentum. In the next chapters we argue that in GFQT it is natural that the width of the momentum distribution for a macroscopic body is inversely proportional to its mass. Then we recover the Newton gravitational law. Namely, if

$$\delta_j = \frac{R}{m_j G'} \quad (j=1,2), \quad C^2 G' = 2G \qquad (5.32)$$

then

$$H(\mathbf{r},\mathbf{q}) = \frac{\mathbf{q}^2}{2m_{12}} - G\frac{m_1 m_2}{r} \qquad (5.33)$$

We conclude that in our approach gravity is simply a dS the correction to the standard nonrelativistic Hamiltonian. This correction is spherically symmetric since, as noted in the beginning of this section, when all corrections of the order of $1/R$ are neglected, the dependence of $\mathbf{D}^2$ on $\mathbf{J}$ disappears.

5.6 Special case: very large m_2

Consider a special case when $m_2 \gg m_1, |\mathbf{q}|$ and we do not assume that particle 1 is nonrelativistic. As noted above, in the c.m. frame of the two-body system $n_1 \approx n_2 \approx n$. Since in this reference frame the vectors $\mathbf{B}_1$ and $\mathbf{B}_2$ are approximately antiparallel and $|\mathbf{B}_1| \approx |\mathbf{B}_2| \approx 2n$, it follows from Eq. (5.7) that

$$\overline{W}_0 = [(w_1+4n^2)^{1/2}+(w_2+4n^2)^{1/2}]^2 \approx [(w_1+4n^2)^{1/2}+w_2^{1/2}]^2 \approx w_2+2w_2^{1/2}(w_1+4n^2)^{1/2} \qquad (5.34)$$

since $w_2 \gg w_1, 4n^2$.

Consider now the calculation of the quantity $F(\delta_1,\delta_2,\varphi)$ in Eq. (5.29). If the quantities δ_i $(i=1,2)$ are inversely proportional to the corresponding masses then $\delta_1 \gg \delta_2$. Now it is clear from Fig. 5.1 that in the sum for $F(\delta_1,\delta_2,\varphi)$ the number of terms approximately equals δ_1 and in almost all of them $\delta(N)=\delta_2$. Hence $F(\delta_1,\delta_2,\varphi) \approx 1-cos\chi$ where $\chi = f(C/(\varphi\delta_2)^{1/2})$. Then, as follows from Eqs. (5.29) and (5.34)

$$\overline{W}^{1/2} \approx w_2^{1/2} + (w_1+4n^2)^{1/2} cos\chi \qquad (5.35)$$

Equation (5.35) is derived neglecting all the corrections of the order of $1/R$ and higher; it particular it is assumed that $k \ll n$. Hence the last term in Eq. (5.35) differs from the first term in Eq. (4.20) only such that φ is replaced by χ. This is a consequence of the fact that the latter has been derived by considering the single-particle wave function and assuming that its wave function contains $exp(-i\varphi n)$ while the former has been derived by considering the wave function in the c.m. frame and assuming that its dependence on the relative momentum variable n contains $exp(-i\chi n)$.

Since $W = 4R^2M^2$ where M is the standard two-body mass operator, it follows from Eq. (5.35) that if m_2 is very large then the mass operator of the two-body problem is fully defined by the energy of *free* particle 1 in the c.m. frame of the two-body system. For example, when $f(x) \approx x$ then by analogy with the derivation of Eq. (5.33) we get that the energy of particle 1 in the c.m. frame is

$$H_{rel}(\mathbf{r}, \mathbf{q}) = (m_1^2 + \mathbf{q}^2)^{1/2}(1 - \frac{Gm_2}{r}) \tag{5.36}$$

and the nonrelativistic expression for this energy is

$$H_{nr}(\mathbf{r}, \mathbf{q}) = \frac{\mathbf{q}^2}{2m_1} - \frac{Gm_1m_2}{r} \tag{5.37}$$

Let us stress that for example Eq. (5.37) is the nonrelativistic energy of *free* particle 1 in the c.m. frame of the two-body system. In standard theory this expression is treated as a result of gravitational interaction of particle 1 with the massive body having the mass m_2. *Hence in our approach gravity is simply a kinematical consequence of dS symmetry.*

By analogy with the single-body case, the internal two-body wave function can be written as $\psi(n, k, \mu)$ where n is the quantum number characterizing the magnitude of the relative dS momentum, k is the quantum number characterizing the magnitude of the relative angular momentum and μ is the quantum number characterizing the z projection of the relative angular momentum. If $m_1 \ll m_2$ then in the c.m. frame the radius-vector of particle 1 is much greater than the radius-vector of particle 2. As noted above, in the c.m. frame $n_1 \approx n_2 \approx n$. Therefore the relative angular momentum approximately equals the angular momentum of particle 1 in the two-body c.m. frame. As a consequence, $\psi(n, k, \mu)$ can be treated as a wave function of particle 1 in the c.m. frame. The only difference between this wave function and the single-particle wave function for the free particle 1 is that in the case $m_1 \ll m_2$ the width of the n-distribution in the c.m. frame equals δ_2, not δ_1 as for the free particle 1. As a consequence, the energy of particle 1 in the c.m. frame is described by Eqs. (5.36) and (5.37).

In view of the analogy between the description of free particle 1 and particle 1 in the two-body c.m. frame, for describing semiclassical values of the dS operators of particle 1 in the c.m. frame one can use the results of Sec. 4.2 and Eq.

(4.20) where φ is replaced by χ. The classical motion of particle 1 in the xy plane such that $J_z > 0$ corresponds to the case $\alpha = -\pi/2$ and $\mu = k$. Then, taking into account that $k \ll n$, it follows from Eq. (4.20) that

$$\begin{aligned}
&B_x = -2nsin\beta, \quad B_y = 2ncos\beta, \quad J_z = 2k, \quad B_z = N_z = \mathbf{J}_\perp = 0 \\
&\mathcal{E} = (w+4n^2)^{1/2}cos\chi, \quad N_x = (w+4n^2)^{1/2}(sin\chi sin\beta - \frac{k}{n}cos\chi cos\beta) \\
&N_y = (w+4n^2)^{1/2}(-sin\chi cos\beta - \frac{k}{n}cos\chi sin\beta)
\end{aligned} \tag{5.38}$$

For describing vectors in the xy plane we will use the following notation. If the vector $\mathbf{A}$ has the components (A_x, A_y) then we will write $\mathbf{A} = (A_x, A_y)$. As in Sec. 4.2, the relation between the momentum $\mathbf{q}$ of particle 1 in the c.m. frame and the vector $\mathbf{B}$ is $\mathbf{q} = \mathbf{B}/2R$, the standard energy E equals $\mathcal{E}/2R$ and the $\perp$ and $||$ components of the vector $\mathbf{N}$ are defined as in Sec. 4.3. Then, as follows from Eq. (5.38), $\mathbf{N}_\perp = -2ERk(cos\beta, sin\beta)/n$. Since $\mathbf{r}_\perp$ is defined such that $\mathbf{N}_\perp = -2E\mathbf{r}_\perp$ and $n = Rq$ where $q = |\mathbf{q}|$ we get that $\mathbf{r}_\perp = k(cos\beta, sin\beta)/q$ and hence the vector $\mathbf{r}$ can be written as $\mathbf{r} = r_{||}(sin\beta, -cos\beta) + |\mathbf{r}_\perp|(cos\beta, \sin\beta)$.

Since we work in units where $\hbar/2 = 1$ then $k = |\mathbf{r}_\perp|q$ and in standard units $J_z = L$ and $|\mathbf{r}_\perp| = L/q$. We now define the angles γ_1 and γ_2 such that $\beta = \pi/2 + \gamma_1$, $sin\gamma_2 = L/qr$ and $cos\gamma_2 = [1-(L/qr)^2]^{1/2}$ where $r = |\mathbf{r}|$. Then the final result for the vectors $\mathbf{q}$ and $\mathbf{r}$ can be writtens as

$$\mathbf{q} = q(1-\frac{L^2}{q^2r^2})^{1/2}(cos\varphi, sin\varphi) + \frac{L}{r}(-sin\varphi, cos\varphi), \quad \mathbf{r} = r(cos\varphi, sin\varphi) \tag{5.39}$$

where $\varphi = \gamma_1 - \gamma_2$ and we assume that $sin\chi > 0$. The standard energy of particle 1 in the c.m. frame is $E = (m^2 + \mathbf{q}^2)^{1/2}cos\chi$ where $m = m_1$ and χ is a fiction of r discussed in the preceding sections.

5.7 Classical equations of motion

Classical equations of motion should follow from quantum theory if the evolution operator is known. By analogy with standard Schrödinger equation one might think that the internal two-body evolution operator is $exp(-iMt)$ where M is the two-body mass operator. However, as discussed in Sec. 1.2, the problem of time in quantum theory has not been solved yet and such an evolution operator is problematic. Nevertheless, if the evolution operator is defined by M then on classical level the two-body mass and the quantities corresponding to operators commuting with M are conserved. In particular, if L is the classical value of J_3 then L is conserved.

In this section we show that classical equations of motion for all standard gravitational two-body problems can be obtained according to the following scheme. We assume that classical values of the *free* two-body mass M and L are conserved.

In the case when $m_1 \ll m_2$ we assume that, according to Eq. (1.2), time is defined such that

$$d\mathbf{r} = \frac{\mathbf{q}}{\epsilon(q)}dt \tag{5.40}$$

where $\epsilon(q) = (m^2 + q^2)^{1/2}$. Then the results are generalized to the case when m_1 and m_2 are comparable to each other. Note that the above conditions fully define the motion; in particular there is no need to involve Lagrange equations, Hamilton equations, Hamilton-Jacobi equations etc.

Consider first the case when $m_1 \ll m_2$. The three classical tests of GR — precession of Mercury's perihelion, gravitational red shift of light and deflection of light by the Sun — can be discussed in this approximation. If $\xi = sin^2\chi$ then, as discussed in the preceding sections, ξ can be written as a series in powers of (r_g/r) where r_g is the gravitational radius of particle 2: $\xi = (r_g/r) + a(r_g/r)^2 + ...$

The consideration of the gravitational red shift of light does not require Eq. (5.40) and equations of motion. In that case it siffices to note that, according to Eq. (5.36), if particle 1 is the photon then in the approximation when $\xi = r_g/r$ its energy in standard units is

$$E = qc(1 - \frac{r_g}{2r}) = qc(1 - \frac{Gm_2}{c^2 r}) \tag{5.41}$$

Consider the case when the photon travels in the radial direction from the Earth surface to the height h. Let R_E be the Earth radius, q_1 be the photon momentum on the Earth surface when $r = R_E$ and q_2 be the photon momentum when the photon is on the height h, i.e. when $r = R_E + h$. The corresponding photon kinetic energies are $E_1 = q_1 c$ and $E_2 = q_2 c$, respectively. Since E is the conserved quantity, it easily follows from Eq. (5.41) that if $h \ll R_E$ then $\Delta E_1 = E_2 - E_1 \approx -E_1 gh/c^2$ where g is the free fall acceleration. Therefore one can formally define the potential energy of the photon near the Earth surface by $U(h) = E_1 gh/c^2$ and we have a full analogy with classical mechanics. From the formal point of view, the result is in agreement with GR and the usual statement is that this effect has been measured in the famous Pound-Rebka experiment. We discuss this question in Sec. 5.8.

Consider now the derivation of equations of motions in the case when $m_1 \ll m_2$. As follows from Eq. (5.39)

$$d\mathbf{r} = dr(cos\varphi, sin\varphi) + rd\varphi(-sin\varphi, cos\varphi) \tag{5.42}$$

Therefore, as follows from Eqs. (5.39) and (5.40), the equations of motion have the form

$$\frac{dr}{dt} = \frac{1}{\epsilon(q)}(q^2 - \frac{L^2}{r^2})^{1/2}, \quad \frac{d\varphi}{dt} = \frac{L}{r^2\epsilon(q)} \tag{5.43}$$

where q as a function of r should be found from the condition that E is a constant of motion. Since $E = \epsilon(q)cos\chi$, we have that

$$q(r)^2 = \frac{E^2}{1 - \xi(r)} - m^2 \tag{5.44}$$

In such problems as deflection of light by the Sun and precession of Mercury's perihelion it suffices to find only the trajectory of particle 1. As follows from Eq. (5.43), the equation defining the trajectory is

$$\frac{d\varphi}{dr}=\frac{L}{r[r^2q(r)^2-L^2]^{1/2}} \tag{5.45}$$

Consider first deflection of light by the Sun. If ρ is the minimal distance between the photon and the Sun then when $r=\rho$ the radial component of the momentum is zero and hence, as follows from Eq. (5.41)

$$q(\rho)=\frac{L}{\rho},\quad E=\frac{L}{\rho}(1-\frac{r_g}{2\rho}) \tag{5.46}$$

Suppose that in Eq. (5.44) a good approximation is when only the terms linear in r_g/r can be taken into account. Then $q(r)^2\approx E^2(1+r_g/r)$ and, as follows from Eqs. (5.45) and (5.46), in first order in r_g/r

$$\frac{d\varphi}{dr}=\frac{\rho}{r}[r^2(1+\frac{r_g}{r}-\frac{r_g}{\rho})-\rho^2]^{-1/2} \tag{5.47}$$

Suppose that in the initial state the y coordinate of the photon was $-\infty$, at the closest distance to the Sun its coordinates are $(x=\rho, y=0)$ and in the final state the y coordinate is $+\infty$. Then, as follows from Eq. (5.47), the total change of the photon angle is

$$\Delta\varphi=2\int_{\rho}^{\infty}\frac{\rho}{r}[r^2(1+\frac{r_g}{r}-\frac{r_g}{\rho})-\rho^2]^{-1/2}dr \tag{5.48}$$

The quantities r_g/ρ and r_g/r are very small and in the main approximation those quantities can be neglected. Then $\Delta\varphi=\pi$ what corresponds to the non-deflected motion along a straight line. In the next approximation in r_g/ρ

$$\Delta\varphi=\pi+\frac{r_g}{\rho} \tag{5.49}$$

This result is discussed in Sec. 5.8.

Consider now the trajectory of particle 1 if its mass $m=m_1$ is arbitrary but such that $m_1\ll m_2$ and the terms quadratic in r_g/r should be taken into account. Then $E/(1-\xi)\approx E(1+\xi+\xi^2)$ and, as follows from Eqs. (5.44) and (5.45)

$$\frac{d\varphi}{dr}=\frac{L}{r}[(E^2-m^2)r^2+E^2r_gr+E^2r_g^2(1+a)-L^2]^{-1/2} \tag{5.50}$$

If $E<m$ then it is clear from this expression that the quantity r can be only in a finite range $[r_1,r_2]$.

For defining the trajectory one can use the fact that

$$\int\frac{dx}{x(-ax^2+bx-c)^{1/2}}=\frac{i}{\sqrt(c)}ln[A(x)+iB(x)]$$

where

$$A(x) = \frac{2}{x}(-ax^2 + bx - c)^{1/2}, \quad B(x) = \frac{i(bx-2c)}{xc^{1/2}}$$

Since $lnz = ln|z| + iarg(z)$, the result of integration of Eq. (5.50) is

$$\varphi(r) = const + \frac{L}{[L^2 - E^2 r_g^2(1+a)]^{1/2}} arcsin[F(r)] \tag{5.51}$$

where the explicit form of the function $F(r)$ is not important for our goal. It follows from this expression that the difference of the angles for consecutive perihelions is

$$\Delta\varphi = \frac{2\pi L}{[L^2 - E^2 r_g^2(1+a)]^{1/2}} \tag{5.52}$$

If $E^2 r_g^2(1+a) \ll L^2$ and particle 1 is nonrelativistic this expression can be written as

$$\Delta\varphi = 2\pi + \frac{4\pi m^2 m_2^2 G^2(1+a)}{L^2} \tag{5.53}$$

and the result of GR is recovered if $a = 1/2$. This result is discussed in Sec. 5.8.

Note that in the three classical tests of GR we need only trajectories, i.e. the knowledge of the functions $r(t)$ and $\varphi(t)$ is not needed. Then it is clear that although for the derivation of Eq. (5.45) we used Eq. (5.40), the only property of this equation needed for defining trajectories is that $d\mathbf{r}$ is proportional to $\mathbf{q}$. However, for defining the functions $r(t)$ and $\varphi(t)$ it is important that $d\mathbf{r}/dt$ is the velocity defined as $\mathbf{q}/\epsilon(q)$.

For example, as follows from Eqs. (5.43) and (5.44) the relation between t and r is

$$t(r) = E\int \frac{dr}{[1-\xi(r)]^{1/2}[(E^2-m^2) + E^2\xi(r)(1+\xi(r)) - L^2/r^2]^{1/2}} \tag{5.54}$$

Taking into account corrections of the order of r_g/r we get

$$t(r) = E\int \frac{(r + r_g/2)dr}{[(E^2-m^2)r^2 + E^2 r_g r + E^2(1+a)r_g^2 - L^2]^{1/2}} \tag{5.55}$$

Let T be the period of rotations; for example it can defined as the time difference between two consecutive perihelions. This quantity can be calculated by analogy with the above calculation of angular precession of the perihelion and the result is

$$T = \frac{\pi E m^2 r_g}{(m^2 - E^2)^{3/2}} \tag{5.56}$$

Suppose that particle 1 is nonrelativistic and define $E_{nr} = m - E$. Then

$$T = T_{nr}(1 - \frac{E_{nr}}{4m}), \quad T_{nr} = \frac{\pi m^3}{(2mE_{nr})^{3/2}} \tag{5.57}$$

where T_{nr} is the nonrelativistic expression for the period. It follows from this expression that the relativistic correction to the period is $2.4 \cdot 10^{-2}s$ for Mercury and $3.9 \cdot 10^{-2}s$ for Earth. In GR the period can be calculated by using the expression for $t(r)$ in this theory (see e.g. Ref. [3]). For Earth this gives an additional correction of $0.6s$. However, at present the comparison between theory and experiment with such an accuracy seems to be impossible.

In standard nonrelativistic theory the acceleration $d^2\mathbf{r}/dt^2$ is directed toward the center and is proportional to $1/r^2$. Let us check whether or not this property is satisfied in the above formalism. As follows from Eq. (5.39)

$$\frac{d^2\mathbf{r}}{dt^2} = [\frac{d^2r}{dt^2} - r(\frac{d\varphi}{dt})^2](cos\varphi, sin\varphi) + [2\frac{dr}{dt}\frac{d\varphi}{dt} + r\frac{d^2\varphi}{dt^2}](-sin\varphi, cos\varphi) \tag{5.58}$$

Therefore $d^2\mathbf{r}/dt^2$ is directed toward $\mathbf{r}$ if $[2(dr/dt)(d\varphi/dt) + rd^2\varphi/dt^2] = 0$. A direct calculation using Eqs. (5.43) and (5.44) gives

$$2\frac{dr}{dt}\frac{d\varphi}{dt} + r\frac{d^2\varphi}{dt^2} = -\frac{LE^2(1+2\xi)}{2\epsilon(q)^3 r}\frac{d\xi}{dt} \tag{5.59}$$

In the nonrelativistic approximation this quantity does equal zero but in the general case it does not.

An analogous calculation gives

$$\frac{d^2r}{dt^2} - r(\frac{d\varphi}{dt})^2 = \frac{(m^2 + L^2/r^2)E^2(1+2\xi)}{2(q^2 - L^2/r^2)^{1/2}\epsilon(q)^3}\frac{d\xi}{dt} \tag{5.60}$$

If $\xi = r_g/r$ then, as follows from Eq. (5.43), in the nonrelativistic approximation this quantity equals $-Gm_2/r^2$. Hence in this approximation we indeed have the standard result $d^2\mathbf{r}/dt^2 = -Gm_2\mathbf{r}/r^3$.

We now do not assume that $m_1 \ll m_2$ but consider only the nonrelativistic approximation. The relative angular momentum $\mathbf{J}$ equals the total angular momentum in the c.m. frame. In this reference frame we have $\mathbf{r}_1 \times \mathbf{p}_1 + \mathbf{r}_2 \times \mathbf{p}_2 = \mathbf{r} \times \mathbf{q}$ where $\mathbf{q} = \mathbf{p}_1$ is the relative momentum and $\mathbf{r} = \mathbf{r}_1 - \mathbf{r}_2$ is the relative position. Therefore, by analogy with the derivation of Eq. (5.39), one can derive the same relations where $\mathbf{q}$ is the relative momentum and $\mathbf{r}$ is the relative position.

As follows from Eqs. (5.8) and (5.33), in the cases of dS antigravity and standard Newtonian gravity the internal two-body nonrelativistic energy can be written as

$$E = \frac{\mathbf{q}^2}{2m_{12}} - \frac{1}{2}m_{12}\xi \tag{5.61}$$

where $\xi = (r/R)^2$ for the dS antigravity and $\xi = r_g/r$ with $r_g = 2G(m_1 + m_2)$ for the Newtonian gravity.

By analogy with the above consideration, for deriving equations of notions one should define time by analogy with Eq. (5.40). In our approach the effects of dS antigravity and standard Newtonian gravity are simply kinematical manifestations of dS symmetry for systems of two free particles. The difference between those cases is only that the quantity χ in the exponent $exp(-i\chi n)$ defining the behavior of the internal two-body wave function depends on r differently in the cases when r is of the order of cosmological distances and much less than those distances. As follows from Eq. (5.40), in the nonrelativistic approximation

$$d\mathbf{r} = d\mathbf{r}_1 - d\mathbf{r}_2 = (\frac{\mathbf{p}_1}{m_1} - \frac{\mathbf{p}_2}{m_2})dt = \frac{\mathbf{q}}{m_{12}}dt$$

Then, by analogy with the derivation of Eq. (5.58), we get

$$\frac{d^2\mathbf{r}}{dt^2} = \frac{\mathbf{r}}{R^2}, \quad m_{12}\frac{d^2\mathbf{r}}{dt^2} = -\frac{Gm_1m_2}{r^3}\mathbf{r} \tag{5.62}$$

for dS antigravity and Newtonian gravity, respectively.

As noted in Secs. 3.6 and 5.1, the first expression in Eq. (5.62) is a consequence of the Hamilton equations for the Hamiltonian (5.8) and it is obvious that the second expression is a consequence of the Hamilton equations for the Hamiltonian (5.33). However, as shown above, those expressions can be derived without involving the Hamilton equations but using only the relation (5.40) for each free particle.

5.8 Discussion of non-Newtonian gravitational effects

General Relativity is a pure classical theory and a common belief is that in the future quantum theory of gravity the results of GR will be recovered in semiclassical approximation. Moreover, any quantum theory of gravity can be tested only on macroscopic level. Hence, the problem is not only to construct quantum theory of gravity but also to understand a correct structure of the position operator on macroscopic level. However, in the literature the latter problem is not discussed because it is tacitly assumed that the position operator on macroscopic level is the same as in standard quantum theory. This is a great extrapolation which should be substantiated.

As argued in Secs. 5.3 and 5.4, standard position operator is not semiclassical on macroscopic level and therefore on this level it should be modified. In our approach gravity is simply a manifestation of dS symmetry on quantum level for systems of free bodies. Then for calculating observable effects one should know how the quantity χ in the exponent $exp(-i\chi n)$ for the internal two-body wave function depends on the distance between the bodies. As argued in Sec. 5.4, if $\xi = sin^2\chi$ then the dependendence $\xi = (r_g/r) + o(r_g/r)$ is reasonable and reproduces standard Newtonian gravity. In this section we consider what our approach can say about the gravitational red shift of light, deflection of light by the Sun and precession of Mercury's perihelion which are treated as three classical tests of GR.

As seen from Earth the precession of Mercury's orbit is measured to be 5600" per century while the contribution of GR is 43" per century. Hence the latter is less than 1% of the total contribution. The main contribution to the total precession arises as a consequence of the fact that Earth is not an inertial reference frame and when the precession is recalculated with respect to the International Celestial Reference System the value of the precession becomes (574.10 ± 0.65)" per century. Celestial mechanics states the gravitational tugs of the other planets contribute (531.63 ± 0.69)" while all other contributions are small. Hence there is a discrepancy of 43" per century and the result of GR gives almost exactly the same value. Although there are different opinions on whether, the contribution of GR fully explains the data or not, in the overwhelming majority of the literature it is accepted that this is the case.

Our result (5.53) is compatible with GR if $\xi = (r_g/r) + (r_g/r)^2/2 + o((r_g/r)^2)$. The result of GR is by a factor of 3/2 greater than the results of several alternative theories of gravity which in our approach can be reproduced if $\xi = (r_g/r) + o((r_g/r)^2)$. Hence the problem of the future quantum theory of gravity is to understand the value of the quadratic correction to ξ.

The result for the gravitational red shift of light given by Eq. (5.41) is in agreement with GR and is treated such that it has been confirmed in the Pound-Rebka experiment. However, the conventional interpretation of this effect has been criticized by L.B. Okun in Ref. [92]. In his opinion, "*a presumed analogy between a photon and a stone*" is wrong. The reason is that "*the energy of the photon and hence its frequency $\omega = E/\hbar$ do not depend on the distance from the gravitational body, because in the static case the gravitational potential does not depend on the time coordinate t. The reader who is not satisfied with this argument may look at Maxwell's equations as given e.g. in section 5.2 of ref. [93]. These equations with time independent metric have solutions with frequencies equal to those of the emitter*". In Ref. [92] the result of the Pound-Rebka experiment is explained such that not the photon looses its kinetic energy but the differences between the atom energy levels on the height h are greater than on the Earth surface and "*As a result of this increase the energy of a photon emitted in a transition of an atom downstairs is not enough to excite a reverse transition upstairs. For the observer upstairs this looks like a redshift of the photon. Therefore for a competent observer the apparent redshift of the photon is a result of the blueshift of the clock.*".

As noted in Ref. [92], "*A naive (but obviously wrong!) way to derive the formula for the redshift is to ascribe to the photon with energy E a mass $m_\gamma = E/c^2$ and to apply to the photon a non-relativistic formula $\Delta E = -m_\gamma \Delta\phi$ treating it like a stone. Then the relative shift of photon energy is $\Delta E/E = -\Delta\phi/c^2$, which coincides with the correct result. But this coincidence cannot justify the absolutely thoughtless application of a nonrelativistic formula to an ultrarelativistic object.*"

However, in our approach no nonrelativistic formulas for the photon have been used and the result $\Delta E_1/E_1 = -gh/c^2$ has been obtained in a fully relativistic approach. As already noted, the only problematic point in deriving this result is that the function $\xi(r)$ is not exactly known. In the framework of our approach a stone and a photon are simply particles with different masses; that is why the stone is nonrelativistic and the photon is ultrarelativistic. Therefore there is no reason to think that in contrast to the stone, the photon will not loose its kinetic energy. At the same time, we believe that Ref. [92] gives strong arguments that energy levels on the Earth surface and on the height H are different.

We believe that the following point in the arguments of Ref. [92] is not quite consistent. A stone, a photon and other particles can be characterized by their energies, momenta and other quantities for which there exist well defined operators. Those quantities might be measured in collisions of those particles with other particles. At the same time, as noted in Secs. 1.2 and 2.6 the notions of "frequency of a photon" or "frequency of a stone" have no physical meaning. If a particle wave function (or, as noted in Sec. 1.2, rather a state vector is a better name) contains $exp[i(px-Et)/\hbar]$ then by analogy with the theory of classical waves one might say that the particle is a wave with the frequency $\omega = E/\hbar$ and the wave length $\lambda = 2\pi\hbar/p$. However, the fact that such defined quantities ω and λ are the real frequencies and wave lengths measured e.g. in spectroscopic experiments needs to be substantiated. Let ω and λ be frequencies and wave lengths measured in experiments with classical waves. Those quantities necessarily involve classical space and time. Then the relation $E = \hbar\omega$ between the energies of particles in classical waves and frequencies of those waves is only an assumption that those different quantities are related in such a way. This relation has been first proposed by Planck for the description of the blackbody radiation and the experimental data indicate that it is valid with a high accuracy. As noted in Sec. 2.6, this relation takes place in Poincare invariant electrodynamics. However, there is no guaranty that this relation is always valid with the absolute accuracy, as the author of Ref. [92] assumes. In spectroscopic experiments not energies and momenta of emitted photons are measured but wave lengths of the radiation obtained as a result of transitions between different energy levels. In particular, there is no experiment confirming that the relation $E = \hbar\omega$ is always exact, e.g. on the Earth surface and on the height h. In summary, the Pound-Rebka experiment cannot be treated as a model-independent confirmation of GR.

Consider now the deflection of light by the Sun. As shown in the preceding section, in the approximation $\xi = r_g/r$ the deflection is described by Eq. (5.49). In the literature this result is usually represented such that if $\theta = \Delta\varphi - \pi$ is the deflection angle then $\theta = (1+\gamma)r_g/\rho$ where γ depends on the theory. Hence the result given by Eq. (5.49) corresponds to $\gamma = 0$. This result was obtained by Einstein in 1911. The well-known historical facts are that in 1915 when Einstein created GR he obtained $\gamma = 1$ and in 1919 this result was confirmed in observations of the full Solar eclipse. Originally the accuracy of measurements was not high but now the quantity γ is measured with a high accuracy in experiments using the Very Long Base Interferometry (VLBI) technique and the result $\gamma = 1$ has been confirmed with the accuracy better than 1%. The result $\gamma = 1$ in GR is a consequence of the fact that the post-Newtonian correction to the metric tensor in the vicinity of the Sun is not zero for both, temporal and spatial components of this tensor. A question arises whether this result can be obtained in the framework of a quantum approach. In the textbook [94], the deflection is treated as a consequence of one-graviton exchange. The author defines the vertices responsible for the interaction of a virtual graviton with a scalar nonrelativistic particle and with a photon and in that case the cross-section of the process described by the one-graviton exchange corresponds to the result with $\gamma = 1$. The problem is that there is no other way of testing the photon-graviton vertex and we believe that it is highly unrealistic that when the photon travels in the y direction from $-\infty$ to $+\infty$, it exchanges only by one virtual graviton with the Sun. Therefore a problem of how to recover the result with $\gamma = 1$ in quantum theory remains open.

In GR it is assumed that in the propagation of light in the interstellar medium the interaction of light with the medium is not significant and the propagation can be described in the framework of geometrical optics. In other words, this approach is similar to what is called Theory A for explaining the redshift (see Chap. 2). However, the density of the Solar atmosphere near the Solar surface is rather high and the assumption that the photon passes this atmosphere practically without interaction with the particles of the atmosphere seems to be problematic.

For example, in Sec. 2.10 we discussed possible mechanisms which do not allow the photon wave function to spread significantly. In particular, a possible mechanism can be such that a photon is first absorbed by an atom and then is reemitted. Suppose that this mechanism plays an important role and photons encounter many atoms on their way. In the period of time when the atom absorbs the photon but does not reemit it yet, the atom acquires an additional acceleration as a result of its effective gravitational interaction with the Sun. Then the absorbed and reemitted photons will have different accelerations and the reemitted photon is expected to have a greater acceleration towards the Sun than the absorbed photon. This effect increases the deflection angle and analogously other mechanisms of interaction of photons with the interstellar matter are expected to increase the deflection angle since the matter moves with an acceleration towards the Sun.

Three classical effects of GR are treated as phenomena where the gravitational field is rather weak. In recent years considerable efforts have been made for investigating binary pulsars where the gravitational field is treated as strong. In contrast to planets, conclusions about masses and radii of pulsars can be made only from models describing their radiation. It is believed that typically pulsars are neutron stars with masses in the range 1.2-1.6 of the solar one and radii of the order of $10km$. In the case of binary pulsars, a typical situation is that the second component of the binary system is not observable (at present the only known case where the both components are pulsars is the binary pulsar J0737-3039).

The most famous case is the binary pulsar PSR B1913+16 discovered by Hulse and Taylor in 1974. A model with eighteen fitted parameters for this binary system has been described in Refs. [95, 96] and references therein. In this model the masses of the pulsar and companion are approximately 1.4 solar masses, the period of rotation around the common center of mass is 7.75 hours, the values of periastron and apastron are 1.1 and 4.8 solar radii, respectively, and the orbital velocity of stars is 450 km/s and 110 km/s at periastron and apastron, respectively. Then relativistic effects are much stronger than in Solar System. For example, the precession of periastron is 4.2 degrees per year.

The most striking effect in the above model is that it predicts that the energy loss due to gravitational radiation can be extracted from the data. As noted in Ref. [95], comparison of the measured and theoretical values requires a small correction for relative acceleration between the solar system and binary pulsar system, projected onto the line of sight. The correction term depends on several rather poorly known quantities, including the distance and proper motion of the pulsar and the radius of the Sun's galactic orbit. However, with best currently available values the agreement between the data and the Einstein quadrupole formula for the gravitational radiation is better than 1%. The rate of decrease of orbital period is 76.5 microseconds per year (i.e. one second per 14000 years).

As noted by the authors of Ref. [95], "*Even with 30 years of observations, only a small portion of the North-South extent of the emission beam has been observed. As a consequence, our model is neither unique nor particularly robust. The North-South symmetry of the model is assumed, not observed, since the line of sight has fallen on the same side of the beam axis throughout these observations. Nevertheless, accumulating data continue to support the principal features noted above.*"

The size of the invisible component is not known. The arguments that this component is a compact object are as follows [97]: "Because the orbit is so close ($1 solar radius$)) and because there is no evidence of an eclipse of the pulsar signal or of mass transfer from the companion, it is generally agreed that the companion is compact. Evolutionary arguments suggest that it is most likely a dead pulsar, while B1913+16 is a recycled pulsar. Thus the orbital motion is very clean, free from tidal or other complicating effects. Furthermore, the data acquisition is clean in the sense that by exploiting the intrinsic stability of the pulsar clock combined with the ability to maintain and transfer atomic time accurately using GPS, the observers can keep track of pulse time-of-arrival with an accuracy of $13\mu s$, despite extended gaps between observing sessions (including a several-year gap in the middle 1990s for an upgrade of the Arecibo radio telescope). The pulsar has shown no evidence of glitches in its pulse period." However, it is not clear whether or not there exist other reasons for substantial energy losses. For example, since the bodies have large velocities and are moving in the interstellar medium, it is not clear whether their interaction with the medium can be neglected.

Nevertheless, the above results are usually treated as a strong indirect confirmation of the existence of gravitational waves. Those results have given a motivation for building powerful facilities where the gravitational waves are expected to be detected directly. However, after more than ten years of observations no unambiguous detections of gravitational waves have been reported.

The discussion in this section shows that the problem of explaining non-Newtonian gravitational effects is very complicated and at present any conclusion about them can be based only on model dependent approaches. So the statements that those effects can be treated as strong confirmations of GR are premature. In any case until the nature of gravity on classical and quantum level is well understood, different approaches should be investigated.

Chapter 6

Why is GFQT more pertinent physical theory than standard one?

6.1 What mathematics is most pertinent for quantum physics?

As noted in Sec. 1.1, several strong arguments indicate that fundamental quantum theory should be based on discrete mathematics. In this chapter we consider an approach when this theory is based on a Galois field. Since the absolute majority of physicists are not familiar with Galois fields, our first goal in this chapter is to convince the reader that the notion of Galois fields is not only very simple and elegant, but also is a natural basis for quantum physics. If a reader wishes to learn Galois fields on a more fundamental level, he or she might start with standard textbooks (see e.g. Ref. [98]).

In view of the present situation in modern quantum physics, a natural question arises why, in spite of big efforts of thousands of highly qualified physicists for many years, the problem of quantum gravity has not been solved yet. We believe that a possible answer is that they did not use the most pertinent mathematics.

For example, the problem of infinities remains probably the most challenging one in standard formulation of quantum theory. As noted by Weinberg [2], *'Disappointingly this problem appeared with even greater severity in the early days of quantum theory, and although greatly ameliorated by subsequent improvements in the theory, it remains with us to the present day'*. The title of Weinberg's paper [99] is "Living with infinities". A desire to have a theory without divergences is probably the main motivation for developing modern theories extending QFT, e.g. loop quantum gravity, noncommutative quantum theory, string theory etc. On the other hand, in theories over Galois fields, infinities cannot exist in principle since any Galois field is finite.

The key ingredient of standard mathematics is the notions of infinitely small and infinitely large. As already noted in Sec. 1.1, in view of the fact that matter is discrete, the notions of standard division and infinitely small can have only a limited applicability. Then we have to acknowledge that fundamental physics cannot be based on continuity, differentiability, geometry, topology etc. As noted in Sec. 1.1, the reason why modern quantum physics is based on these notions is probably a consequence of the fact that discrete mathematics still is not a part of standard physics education.

The notion of infinitely large is based on our belief that *in principle* we can operate with any large numbers. In standard mathematics this belief is formalized in terms of axioms about infinite sets (e.g. Zorn's lemma or Zermelo's axiom of choice) which are accepted without proof. The belief that these axioms are correct is based on the fact that sciences using standard mathematics (physics, chemistry etc.) describe nature with a very high accuracy. It is believed that this is much more important than the fact that, as follows from Gödel's incompleteness theorems, standard mathematics is not a self-consistent theory.

Standard mathematics contains statements which seem to be counterintuitive. For example, the interval $(0,1)$ has the same cardinality as $(-\infty,\infty)$. Another example is that the function tgx gives a one-to-one relation between the intervals $(-\pi/2,\pi/2)$ and $(-\infty,\infty)$. Therefore one can say that a part has the same number of elements as a whole. One might think that this contradicts common sense but in standard mathematics the above facts are not treated as contradicting.

While Gödel's works on the incompleteness theorems are written in highly technical terms of mathematical logics, the fact that standard mathematics has foundational problems is clear from the philosophy of quantum theory. Indeed in this philosophy there should be no statements accepted without proof (and based only on belief that they are correct); only those statements should be treated as physical, which can be experimentally verified, at least in principle. For example, the first incompleteness theorem says that not all facts about natural numbers can be proved. However, from the philosophy of quantum theory this seems to be clear because we cannot verify that $a+b=b+a$ for any numbers a and b.

Suppose we wish to verify that 100+200=200+100. In the spirit of quantum theory it is insufficient to just say that 100+200=300 and 200+100=300. We should describe an experiment where these relations can be verified. In particular, we should specify whether we have enough resources to represent the numbers 100, 200 and 300. We believe the following observation is very important: although standard mathematics is a part of our everyday life, people typically do not realize that *standard mathematics is implicitly based on the assumption that one can have any desirable amount of resources.*

Suppose, however that our world is finite. Then the amount of resources cannot be infinite. In particular, it is impossible in principle to build a computer operating with any number of bits. In this scenario it is natural to assume that there exists a fundamental number p such that all calculations can be performed only modulo p. Then it is natural to consider a quantum theory over a Galois field with the characteristic p. Since any Galois field is finite, the fact that arithmetic in this field is correct can be verified (at least in principle) by using a finite amount of resources.

Let us look at mathematics from the point of view of the famous Kronecker expression: "*God made the natural numbers, all else is the work of man*". Indeed, the natural numbers 0, 1, 2... have a clear physical meaning. However only two operations are always possible in the set of natural numbers: addition and multiplication. In order to make addition reversible, we introduce negative integers -1, -2 etc. Then, instead of the set of natural numbers we can work with the ring of integers where three operations are always possible: addition, subtraction and multiplication. However, the negative numbers do not have a direct physical meaning (we cannot say, for example, "I have minus two apples"). Their only role is to make addition reversible.

The next step is the transition to the field of rational numbers in which all four operations except division by zero are possible. However, as noted above, division has only a limited meaning.

In mathematics the notion of linear space is widely used, and such important notions as the basis and dimension are meaningful only if the space is considered over a field or body. Therefore if we start from natural numbers and wish to have a field, then we have to introduce negative and rational numbers. However, if, instead of all natural numbers, we consider only p numbers 0, 1, 2, ... $p-1$ where p is prime, then we can easily construct a field without adding any new elements. This construction, called Galois field, contains nothing that could prevent its understanding even by pupils of elementary schools.

Let us denote the set of numbers 0, 1, 2,...$p-1$ as F_p. Define addition and multiplication as usual but take the final result modulo p. For simplicity, let us consider the case $p = 5$. Then F_5 is the set 0, 1, 2, 3, 4. Then $1+2=3$ and $1+3=4$ as usual, but $2+3=0$, $3+4=2$ etc. Analogously, $1 \cdot 2 = 2$, $2 \cdot 2 = 4$, but $2 \cdot 3 = 1$, $3 \cdot 4 = 2$ etc. By definition, the element $y \in F_p$ is called opposite to $x \in F_p$ and is denoted as $-x$ if $x + y = 0$ in F_p. For example, in F_5 we have -2=3, -4=1 etc. Analogously $y \in F_p$ is called inverse to $x \in F_p$ and is denoted as $1/x$ if $xy = 1$ in F_p. For example, in F_5 we have 1/2=3, 1/4=4 etc. It is easy to see that addition is reversible for any natural $p > 0$ but for making multiplication reversible we should choose p to be a prime. Otherwise the product of two nonzero elements may be zero modulo p. If p is chosen to be a prime then indeed F_p becomes a field without introducing any new objects (like negative numbers or fractions). For example, in this field each element can obviously be treated as positive and negative *simultaneously*!

The above example with division might also be an indication that, in the spirit of Ref. [100], the ultimate quantum theory will be based even not on a Galois field but on a finite ring (this observation was pointed out to me by Metod Saniga).

One might say: well, this is beautiful but impractical since in physics and everyday life 2+3 is always 5 but not 0. Let us suppose, however that fundamental physics is described not by "usual mathematics" but by "mathematics modulo p" where p is a very large number. Then, operating with numbers which are much less than p we will not notice this p, at least if we only add and multiply. We will feel a difference between "usual mathematics" and "mathematics modulo p" only while operating with numbers comparable to p.

We can easily extend the correspondence between F_p and the ring of integers Z in such a way that subtraction will also be included. To make it clearer we note the following. Since the field F_p is cyclic (adding 1 successively, we will obtain 0 eventually), it is convenient to visually depict its elements by the points of a circle of the radius $p/2\pi$ on the plane (x, y). In Fig. 6.1 only a part of the circle near the origin is depicted. Then the distance between neighboring elements of the field is

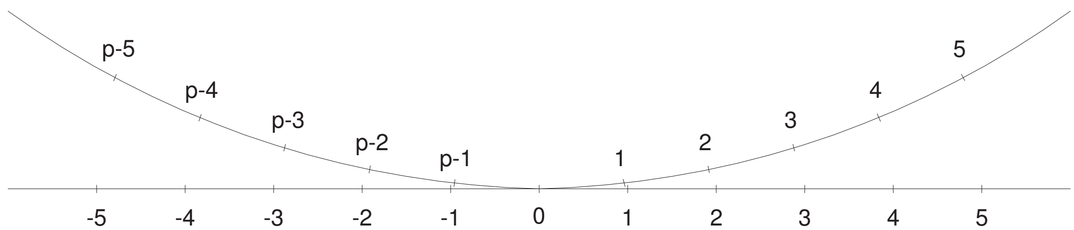

Figure 6.1: Relation between F_p and the ring of integers

equal to unity, and the elements 0, 1, 2,... are situated on the circle counterclockwise. At the same time we depict the elements of Z as usual such that each element $z \in Z$ is depicted by a point with the coordinates $(z, 0)$. We can denote the elements of F_p not only as 0, 1,... $p-1$ but also as 0, ± 1, ± 2,,...$\pm(p-1)/2$, and such a set is called the set of minimal residues. Let f be a map from F_p to Z, such that the element $f(a) \in Z$ corresponding to the minimal residue a has the same notation as a but is considered as the element of Z. Denote $C(p) = p^{1/(lnp)^{1/2}}$ and let U_0 be the set of elements $a \in F_p$ such that $|f(a)| < C(p)$. Then if $a_1, a_2, ...a_n \in U_0$ and n_1, n_2 are such natural numbers that

$$n_1 < (p-1)/2C(p),\ n_2 < ln((p-1)/2)/(lnp)^{1/2} \tag{6.1}$$

then

$$f(a_1 \pm a_2 \pm ...a_n) = f(a_1) \pm f(a_2) \pm ...f(a_n)$$

if $n \leq n_1$ and

$$f(a_1a_2...a_n) = f(a_1)f(a_2)...f(a_n)$$

if $n \leq n_2$. Thus though f is not a homomorphism of rings F_p and Z, but if p is sufficiently large, then for a sufficiently large number of elements of U_0 the addition, subtraction and multiplication are performed according to the same rules as for elements $z \in Z$ such that $|z| < C(p)$. Therefore f can be treated as a local isomorphism of rings F_p and Z.

The above discussion has a well-known historical analogy. For many years people believed that our Earth was flat and infinite, and only after a long period of time they realized that it was finite and had a curvature. It is difficult to notice the curvature when we deal only with distances much less than the radius of the curvature R. Analogously one might think that the set of numbers describing physics has a curvature defined by a very large number p but we do not notice it when we deal only with numbers much less than p. This number should be treated as a fundamental constant describing laws of physics in our World.

One might argue that introducing a new fundamental constant is not justified. However, the history of physics tells us that new theories arise when a parameter, which in the old theory was treated as infinitely small or infinitely large, becomes finite. For example, from the point of view of nonrelativistic physics, the velocity of light c is infinitely large but in relativistic physics it is finite. Analogously, from the point of view of classical theory, the Planck constant $\hbar$ is infinitely small but in quantum theory it is finite. Therefore it is natural to think that in the future quantum physics the quantity p will be not infinitely large but finite.

Let us note that even for elements from U_0 the result of division in the field F_p differs generally speaking, from the corresponding result in the field of rational number Q. For example the element $1/2$ in F_p is a very large number $(p+1)/2$. For this reason one might think that physics based on Galois fields has nothing to do with reality. We will see in the subsequent section that this is not so since the spaces describing quantum systems are projective. It is also clear that in general the meaning of square root in F_p is not the same as in Q. For example, even if $\sqrt{2}$ in F_p exists, it is a very large number of the order of at least $p^{1/2}$. Another obvious fact is that GFQT cannot involve exponents and trigonometric functions since they are represented by infinite sums. Therefore a direct correspondence between wave functions in GFQT and standard theory can exist only for rational functions.

By analogy with the field of complex numbers, we can consider a set F_{p^2} of p^2 elements $a+bi$ where $a, b \in F_p$ and i is a formal element such that $i^2=-1$. The question arises whether F_{p^2} is a field, i.e. we can define all the four operations except division by zero. The definition of addition, subtraction and multiplication in F_{p^2} is obvious and, by analogy with the field of complex numbers, one could define division as $1/(a+bi) = a/(a^2+b^2) - ib/(a^2+b^2)$. This definition can be meaningful only if $a^2+b^2 \neq 0$ in F_p for any $a, b \in F_p$ i.e. a^2+b^2 is not divisible by p. Therefore the definition is meaningful only if p *cannot* be represented as a sum of two squares and is meaningless otherwise. We will not consider the case $p=2$ and therefore p is necessarily odd. Then we have two possibilities: the value of $p\,(mod\,4)$ is either 1 or 3. The well-known result of number theory (see e.g. the textbooks [98]) is that a prime number p can be represented as a sum of two squares only in the former case and cannot in the latter one. Therefore the above construction of the field F_{p^2} is correct only if $p\,(mod\,4)=3$. By analogy with the above correspondence between F_p and Z, we can define a set U in F_{p^2} such that $a+bi \in U$ if $a \in U_0$ and $b \in U_0$. Then if $f(a+bi) = f(a) + f(b)i$, f is a local homomorphism between F_{p^2} and $Z + Zi$.

In general, it is possible to consider linear spaces over any fields. Therefore a question arises what Galois field should be used in GFQT. It is known (see e.g. Ref. [98]) that any Galois field can contain only p^n elements where p is prime and n is natural. Moreover, the numbers p and n define the Galois field up to isomorphism. It is natural to require that there should exist a correspondence between any new theory and the old one, i.e. at some conditions the both theories should give close predictions. In particular, there should exist a large number of quantum states for which the probabilistic interpretation is valid.

In view of the above discussion, the number p should necessarily be very large and the problem is to understand whether there exist deep reasons for choosing a particular value of p, whether this is simply an accident that our world has been created with some value of p, whether the number p is dynamical, i.e. depends on the current state of the world etc. For example, as noted above, the number p defines the existing amount of resources. There are models (see e.g. Ref. [34]) where our world is only a part of the Universe and the amount of resources in the world is not constant.

In any case, if we accept that p is a universal parameter defining what Galois field describes nature (at the present stage of the world or always) then the problem arises what the value of n is. Since we treat GFQT as a more general theory than standard one, it is desirable not to postulate that GFQT is based on F_{p^2} (with $p = 3\ (mod\, 4)$) because standard theory is based on complex numbers but vice versa, explain the fact that standard theory is based on complex numbers since GFQT is based on F_{p^2}. Therefore we should find a motivation for the choice of F_{p^2} with $p = 3\ (mod\, 4)$. Arguments in favor of such a choice are discussed in Refs. [40, 42, 43] and in Secs. 6.3 and 8.9.

6.2 Correspondence between GFQT and standard theory

For any new theory there should exist a correspondence principle that at some conditions this theory and standard well tested one should give close predictions. Well-known examples are that classical nonrelativistic theory can be treated as a special case of relativistic theory in the formal limit $c \to \infty$ and a special case of quantum mechanics in the formal limit $\hbar \to 0$. Analogously, Poincare invariant theory is a special case of dS or AdS invariant theories in the formal limit $R \to \infty$. We treat standard quantum theory as a special case of GFQT in the formal limit $p \to \infty$. Therefore a question arises which formulation of standard theory is most suitable for its generalization to GFQT.

A known historical fact is that quantum mechanics has been originally proposed by Heisenberg and Schrödinger in two forms which seemed fully incompatible with each other. While in the Heisenberg operator (matrix) formulation quantum states are described by infinite columns and operators — by infinite matrices, in the Schrödinger wave formulations the states are described by functions and operators — by differential operators. It has been shown later by Born, von Neumann, Dirac and others that the both formulations are mathematically equivalent. In addition, the path integral approach has been developed.

In the spirit of the wave or path integral approach one might try to replace classical space-time by a finite lattice which may even not be a field. In that case the problem arises what the natural quantum of space-time is and some of physical quantities should necessarily have the field structure. A detailed discussion of this approach can be found in Ref. [101] and references therein. However, as argued in Sect. 1.2, fundamental physical theory should not be based on space-time.

An approach for constructing a quantum theory over a Galois field similar to that proposed in our Refs. [40, 35] and subsequent publications has been discussed in Ref. [102] and references therein.

In the literature there have discussed approaches where quantum theory is based on quaternions or p-adic fields (see e.g. Ref. [103] and references therein). In those approaches infinity still exists and so a problem remains whether or not it is possible to construct quantum theory without divergensies. An approach similar to that proposed in our Refs.

We treat GFQT as a version of the matrix formulation when complex numbers are replaced by elements of a Galois field. We will see below that in that case the columns and matrices are automatically truncated in a certain way, and therefore the theory becomes finite-dimensional (and even finite since any Galois field is finite).

In conventional quantum theory the state of a system is described by a vector $\tilde{x}$ from a separable Hilbert space H. We will use a "tilde" to denote elements of Hilbert spaces and complex numbers while elements of linear spaces over a Galois field and elements of the field will be denoted without a "tilde".

Let $(\tilde{e}_1, \tilde{e}_2, ...)$ be a basis in H. This means that $\tilde{x}$ can be represented as

$$\tilde{x} = \tilde{c}_1\tilde{e}_1 + \tilde{c}_2\tilde{e}_2 + ... \tag{6.2}$$

where $(\tilde{c}_1, \tilde{c}_2, ...)$ are complex numbers. It is assumed that there exists a complete set of commuting selfadjoint operators $(\tilde{A}_1, \tilde{A}_2, ...)$ in H such that each $\tilde{e}_i$ is the eigenvector of all these operators: $\tilde{A}_j\tilde{e}_i = \tilde{\lambda}_{ji}\tilde{e}_i$. Then the elements $(\tilde{e}_1, \tilde{e}_2, ...)$ are mutually orthogonal: $(\tilde{e}_i, \tilde{e}_j) = 0$ if $i \neq j$ where (...,...) is the scalar product in H. In that case the coefficients can be calculated as

$$\tilde{c}_i = \frac{(\tilde{e}_i, \tilde{x})}{(\tilde{e}_i, \tilde{e}_i)} \tag{6.3}$$

Their meaning is that $|\tilde{c}_i|^2(\tilde{e}_i, \tilde{e}_i)/(\tilde{x}, \tilde{x})$ represents the probability to find $\tilde{x}$ in the state $\tilde{e}_i$. In particular, when $\tilde{x}$ and the basis elements are normalized to one, the probability equals $|\tilde{c}_i|^2$.

Let us note that the Hilbert space contains a big redundancy of elements, and we do not need to know all of them. Indeed, with any desired accuracy we can approximate each $\tilde{x} \in H$ by a finite linear combination

$$\tilde{x} = \tilde{c}_1\tilde{e}_1 + \tilde{c}_2\tilde{e}_2 + ...\tilde{c}_n\tilde{e}_n \tag{6.4}$$

where $(\tilde{c}_1, \tilde{c}_2, ...\tilde{c}_n)$ are rational complex numbers. This is a consequence of the known fact that the set of elements given by Eq. (6.4) is dense in H. In turn, this set is redundant too. Indeed, we can use the fact that Hilbert spaces in quantum theory are projective: ψ and $c\psi$ represent the same physical state. Then we can multiply both parts of Eq. (6.4) by a common denominator of the numbers $(\tilde{c}_1, \tilde{c}_2, ...\tilde{c}_n)$. As a result, we can always assume that in Eq. (6.4) $\tilde{c}_j = \tilde{a}_j + i\tilde{b}_j$ where $\tilde{a}_j$ and $\tilde{b}_j$ are integers.

The meaning of the fact that Hilbert spaces in quantum theory are projective is very clear. The matter is that not the probability itself but the relative probabilities of different measurement outcomes have a physical meaning. We believe, the notion of probability is a good illustration of the Kronecker expression about natural numbers (see Sect. 6.1). Indeed, this notion arises as follows. Suppose that conducting experiment N times we have seen the first event n_1 times, the second event n_2 times etc. such that $n_1 + n_2 + ... = N$. We define the quantities $w_i(N) = n_i/N$ (these quantities depend on N) and $w_i = lim\, w_i(N)$ when $N \to \infty$. Then w_i is called the probability of the ith event. We see that all the information about the experiment is given by a set of natural numbers, and in real life all those numbers are finite. However, in order to define probabilities, people introduce additionally the notion of rational numbers and the notion of limit. Another example is the notion of mean value. Suppose we measure a physical quantity such that in the first event its value is q_1, in the second event - q_2 etc. Then the mean value of this quantity is defined as $(q_1n_1 + q_2n_2 + ...)/N$ if N is very large. Therefore, even if all the q_i are integers, the mean value might be not an integer. We again see that rational numbers arise only as a consequence of our convention on how the results of experiments should be interpreted.

The Hilbert space is an example of a linear space over the field of complex numbers. Roughly speaking this means that one can multiply the elements of the space by the elements of the field and use the properties $\tilde{a}(\tilde{b}\tilde{x}) = (\tilde{a}\tilde{b})\tilde{x}$ and $\tilde{a}(\tilde{b}\tilde{x}+\tilde{c}\tilde{y}) = \tilde{a}\tilde{b}\tilde{x} + \tilde{a}\tilde{c}\tilde{y}$ where $\tilde{a}, \tilde{b}, \tilde{c}$ are complex numbers and $\tilde{x}, \tilde{y}$ are elements of the space. The fact that complex numbers form a field is important for such notions as linear dependence and the dimension of spaces over complex numbers.

By analogy with conventional quantum theory, we require that in GFQT linear spaces V over F_{p^2}, used for describing physical states, are supplied by a scalar product (...,...) such that for any $x, y \in V$ and $a \in F_{p^2}$, (x, y) is an element of F_{p^2} and the following properties are satisfied:

$$(x, y) = \overline{(y, x)}, \quad (ax, y) = \bar{a}(x, y), \quad (x, ay) = a(x, y) \tag{6.5}$$

We will always consider only finite dimensional spaces V over F_{p^2}. Let $(e_1, e_2, ...e_N)$ be a basis in such a space. Consider subsets in V of the form $x = c_1e_1 + c_2e_2 + ...c_ne_n$ where for any i, j

$$c_i \in U, \quad (e_i, e_j) \in U \tag{6.6}$$

On the other hand, as noted above, in conventional quantum theory we can describe quantum states by subsets of the form Eq. (6.4). If n is much less than p,

$$f(c_i) = \tilde{c}_i, \quad f((e_i, e_j)) = (\tilde{e}_i, \tilde{e}_j) \tag{6.7}$$

then we have the correspondence between the description of physical states in projective spaces over F_{p^2} on one hand and projective Hilbert spaces on the other. This means that if p is very large then for a large number of elements from V, linear combinations with the coefficients belonging to U and scalar products look in the same way as for the elements from a corresponding subset in the Hilbert space.

In the general case a scalar product in V does not define any positive definite metric and thus there is no probabilistic interpretation for all the elements from V. In particular, $(e, e) = 0$ does not necessarily imply that $e = 0$. However, the probabilistic interpretation exists for such a subset in V that the conditions (6.7) are satisfied. Roughly speaking this means that for elements $c_1e_1 + ...c_ne_n$ such that $(e_i, e_i), c_i\bar{c}_i \ll p$, $f((e_i, e_i)) > 0$ and $c_i\bar{c}_i > 0$ for all $i = 1, ...n$, the probabilistic interpretation is valid. It is also possible to explicitly construct a basis $(e_1, ...e_N)$ such that $(e_j, e_k) = 0$ for $j \neq k$ and $(e_j, e_j) \neq 0$ for all j (see the subsequent section and Chap. 8). Then $x = c_1e_1 + ...c_Ne_N$ $(c_j \in F_{p^2})$ and the coefficients are uniquely defined by $c_j = (e_j, x)/(e_j, e_j)$.

As usual, if A_1 and A_2 are linear operators in V such that

$$(A_1x, y) = (x, A_2y) \quad \forall x, y \in V \tag{6.8}$$

they are said to be conjugated: $A_2 = A_1^*$. It is easy to see that $A_1^{**} = A_1$ and thus $A_2^* = A_1$. If $A = A^*$ then the operator A is said to be Hermitian.

If $(e, e) \neq 0$, $Ae = ae$, $a \in F_{p^2}$, and $A^* = A$, then it is obvious that $a \in F_p$. In the subsequent section (see also Refs. [40, 42]) we will see that there also exist situations when a Hermitian operator has eigenvectors e such that $(e, e) = 0$ and the corresponding eigenvalue is pure imaginary.

Let now $(A_1, ...A_k)$ be a set of Hermitian commuting operators in V, and $(e_1, ...e_N)$ be a basis in V with the properties described above, such that $A_je_i = \lambda_{ji}e_i$. Further, let $(\tilde{A}_1, ...\tilde{A}_k)$ be a set of Hermitian commuting operators in some Hilbert space H, and $(\tilde{e}_1, \tilde{e}_2, ...)$ be some basis in H such that $\tilde{A}_je_i = \tilde{\lambda}_{ji}\tilde{e}_i$. Consider a subset $c_1e_1 + c_2e_2 + ...c_ne_n$ in V such that, in addition to the conditions (6.7), the elements e_i are the eigenvectors of the operators A_j with λ_{ji} belonging to U and such that $f(\lambda_{ji}) = \tilde{\lambda}_{ji}$. Then the action of the operators on such elements have the same form as the action of corresponding operators on the subsets of elements in Hilbert spaces discussed above.

Summarizing this discussion, we conclude that if p is large then there exists a correspondence between the description of physical states on the language of Hilbert spaces and self-adjoint operators in them on one hand, and on the language of linear spaces over F_{p^2} and Hermitian operators in them on the other.

The field of complex numbers is algebraically closed (see standard textbooks on modern algebra, e.g. Ref. [98]). This implies that any equation of the *nth* order in this field always has n solutions. This is not, generally speaking, the case for the field F_{p^2}. As a consequence, not every linear operator in the finite-dimensional space over F_{p^2} has an eigenvector (because the characteristic equation may have no solution in this field). One can define a field of characteristic p which is algebraically closed and contains F_{p^2}. However such a field will necessarily be infinite and we will not use it. We will see in this chapter and Chap. 8 that uncloseness of the field F_{p^2} does not prevent one from constructing physically meaningful representations describing elementary particles in GFQT.

In physics one usually considers Lie algebras over R and their representations by Hermitian operators in Hilbert spaces. It is clear that analogs of such representations in our case are representations of Lie algebras over F_p by Hermitian operators in spaces over F_{p^2}. Representations in spaces over a field of nonzero characteristics are called modular representations. There exists a wide literature devoted to such representations; detailed references can be found for example in Ref. [104] (see also Ref. [40]). In particular, it has been shown by Zassenhaus [105] that all modular IRs are finite-dimensional and many papers have dealt with the maximum dimension of such representations. At the same time, it is worth noting that usually mathematicians consider only representations over an algebraically closed field.

From the previous, it is natural to expect that the correspondence between ordinary and modular representations of two Lie algebras over R and F_p, respectively, can be obtained if the structure constants of the Lie algebra over F_p - c_{kl}^j, and the structure constants of the Lie algebra over R - $\tilde{c}_{kl}^j$, are such that $f(c_{kl}^j) = \tilde{c}_{kl}^j$ (the Chevalley basis [106]), and all the c_{kl}^j belong to U_0. In Refs. [40, 35, 107] modular analogs of IRs of su(2), sp(2), so(2,3), so(1,4) algebras and the osp(1,4) superalgebra have been considered. Also modular representations describing strings have been briefly mentioned. In all these cases the quantities $\tilde{c}_{kl}^j$ take only the values $0, \pm 1, \pm 2$ and the above correspondence does take place.

It is obvious that since all physical quantities in GFQT are discrete, this theory cannot involve any dimensionful quantities and any operators having the continuous spectrum. We have seen in the preceding chapters than the so(1,4) invariant theory is dimensionless and it is possible to choose a basis such that all the operators have only discrete spectrum. As shown in Chap. 8, the same is true for the so(2,3) invariant theories. For this reason one might expect that those theories are natural candidates for their generalization to GFQT. This means that symmetry is defined by the commutation relations (4.1) which are now considered not in standard Hilbert spaces but in spaces over F_{p^2}. We will see in this chapter that there exists a correspondence in the above sense between modular IRs of the finite field analog of the so(1,4) algebra and IRs of the standard so(1,4) algebra and in Chap. 8 the same will be shown for the so(2,3) algebra. At the same time, there is no natural generalization of the Poincare invariant theory to GFQT.

Since the main problems of QFT originate from the fact that local fields interact at the same point, the idea of all modern theories aiming to improve QFT is to replace the interaction at a point by an interaction in some small space-time region. From this point of view, one could say that those theories involve a fundamental length, explicitly or implicitly. Since GFQT is a fully discrete theory, one might wonder whether it could be treated as a version of quantum theory with a fundamental length. Although in GFQT all physical quantities are dimensionless and take values in a Galois field, on a qualitative level GFQT can be thought to be a theory with the fundamental length in the following sense. The maximum value of the angular momentum in GFQT cannot exceed the characteristic of the Galois field p. Therefore the Poincare momentum cannot exceed p/R. This can be interpreted in such a way that the fundamental length in GFQT is of the order of R/p.

One might wonder how continuous transformations (e.g. time evolution or rotations) can be described in the framework of GFQT. A general remark is that if theory $\mathcal{B}$ is a generalization of theory $\mathcal{A}$ then the relation between them is not always straightforward. For example, quantum mechanics is a generalization of classical mechanics, but in quantum mechanics the experiment outcome cannot be predicted unambiguously, a particle cannot be always localized etc. As noted in Sec. 1.2, even in the framework of standard quantum theory, time evolution is well-defined only on macroscopic level. Suppose that this is the case and the Hamiltonian H_1 in standard theory is a good approximation for the Hamiltonian H in GFQT. Then one might think that $exp(-iH_1t)$ is a good approximation for $exp(-iHt)$. However, such a straightforward conclusion is problematic for the following reasons. First, there can be no continuous parameters in GFQT. Second, even if t is somehow discretized, it is not clear how the transformation $exp(-iHt)$ should be implemented in practice. On macroscopic level the quantity Ht is very large and therefore the Taylor series for $exp(-iHt)$ contains a large number of terms which should be known with a high accuracy. On the other hand, one can notice that for computing $exp(-iHt)$ it is sufficient to know Ht only modulo 2π but in this case the question about the accuracy for π arises. We see that a direct correspondence between the standard quantum theory and GFQT exists only on the level of Lie algebras but not on the level of Lie groups.

6.3 Modular IRs of dS algebra and spectrum of dS Hamiltonian

Consider modular analogs of IRs constructed in Sec. 4.1. We noted that the basis elements of this IR are e_{nkl} where at a fixed value of n, $k = 0, 1, ...n$ and $l = 0, 1, ...2k$. In standard case, IR is infinite-dimensional since n can be zero or any natural number. A modular analog of this IR can be only finite-dimensional. The basis of the modular IR is again e_{nkl} where at a fixed value of n the numbers k and l are in the same range as above. The operators of such IR can be described by the same expressions as in Eqs. (4.9-4.14) but now those expressions should be understood as relations in a space over F_{p^2}. However, the quantity n can now be only in the range $0, 1, ...N$ where N can be found from the condition that the algebra of operators described by Eqs. (4.9) and (4.10) should be closed. It follows from these expressions, that this is the case if $w + (2N + 3)^2 = 0$ in F_p and $N + k + 2 < p$. Therefore we have to show that such N does exist.

In the modular case w cannot be written as $w = \mu^2$ with $\mu \in F_p$ since the equality $a^2 + b^2 = 0$ in F_p is not possible if $p = 3\ (mod\, 4)$. In terminology of number theory, this means that w is a quadratic nonresidue. Since -1 also is a quadratic nonresidue if $p = 3\ (mod\, 4)$, w can be written as $w = -\tilde{\mu}^2$ where $\tilde{\mu} \in F_p$ and for $\tilde{\mu}$ obviously two solutions are possible. Then N should satisfy one of the conditions $N + 3 = \pm\tilde{\mu}$ and one should choose that with the lesser value of N. Let us assume that both, $\tilde{\mu}$ and $-\tilde{\mu}$ are represented by $0, 1, ...(p - 1)$. Then if $\tilde{\mu}$ is odd, $-\tilde{\mu} = p - \tilde{\mu}$ is even and *vice versa*. We choose the odd number as $\tilde{\mu}$. Then the two solutions are $N_1 = (\tilde{\mu} - 3)/2$ and $N_2 = p - (\tilde{\mu} + 3)/2$. Since $N_1 < N_2$, we choose $N = (\tilde{\mu} - 3)/2$. In particular, this quantity satisfies the condition $N \leq (p - 5)/2$. Since $k \leq N$, the condition $N + k + 2 < p$ is satisfied and the existence of N is proved. In any realistic scenario, w is such that $w \ll p$ even for macroscopic bodies. Therefore the quantity N should be at least of the order of $p^{1/2}$. The dimension of IR is

$$Dim = \sum_{n=0}^{N} \sum_{k=0}^{n} (2k + 1) = (N + 1)(\frac{1}{3}N^2 + \frac{7}{6}N + 1) \qquad (6.9)$$

and therefore Dim is at least of the order of $p^{3/2}$.

The relative probabilities are defined by $||c(n, k, l)e_{nkl}||^2$. In standard theory the basis states and wave functions can be normalized to one such that the normalization condition is $\sum_{nkl} |\tilde{c}(n, k, l)|^2 = 1$. Since the values $\tilde{c}(n, k, l)$ can be arbitrarily small, wave functions can have an arbitrary carrier belonging to $[0, \infty)$. However, in GFQT the quantities $|c(n, k, l)|^2$ and $||e_{nkl}||^2$ belong to F_p. Roughly speaking, this means that if they are not zero then they are greater or equal than one. Since for probabilistic interpretation we should have that $\sum_{nkl} ||c(n, k, l)e_{nkl}||^2 \ll p$, the probabilistic interpretation may take place only if $c(n, k, l) = 0$ for $n > n_{max}$, $n_{max} \ll N$. That is why in Chap. 4 we discussed only wave functions having the carrier in the range $[n_{min}, n_{max}]$.

As follows from the spectral theorem for selfadjoint operators in Hilbert spaces, any selfadjoint operator A is fully decomposable, i.e. it is always possible to find a basis, such that all the basis elements are eigenvectors (or generalized eigenvectors) of A. As noted in Sect. 6.2, in GFQT this is not necessarily the case since the field F_{p^2} is not algebraically closed. However, it can be shown [98] that for any equation of the Nth order, it is possible to extend the field such that the equation will have N solutions. A question arises what is the minimum extension of F_{p^2}, which guarantees that all the operators $(\mathcal{E}, \mathbf{N}, \mathbf{B}, \mathbf{J})$ are fully decomposable.

The operators $(\mathbf{B}, \mathbf{J})$ describe a representation of the so(4) = su(2)×su(2) subalgebra. It is easy to show (see also Chap. 8) that the representation operators of the su(2) algebra are fully decomposable in the field F_{p^2}. Therefore it is sufficient to investigate the operators $(\mathcal{E}, \mathbf{N})$. They represent components of the so(4) vector operator $M^{0\nu}$ $(\nu = 1, 2, 3, 4)$ and therefore it is sufficient to investigate the dS energy operator $\mathcal{E}$, which with our choice of the basis has a rather simpler form (see Eqs. (4.9) and (4.13)). This operator acts nontrivially only over the variable n and its nonzero matrix elements are given by

$$\mathcal{E}_{n-1,n} = \frac{n+1+k}{2(n+1)}[w + (2n+1)^2] \quad \mathcal{E}_{n+1,n} = \frac{n+1-k}{2(n+1)} \tag{6.10}$$

Therefore, for a fixed value of k it is possible to consider the action of $\mathcal{E}$ in the subspace with the basis elements e_{nkl} $(n = k, k+1, ...N)$.

Let $A(\lambda)$ be the matrix of the operator $\mathcal{E} - \lambda$ such that $A(\lambda)_{qr} = \mathcal{E}_{q+k,r+k} - \lambda\delta_{qr}$. We use $\Delta_q^r(\lambda)$ to denote the determinant of the matrix obtained from $A(\lambda)$ by taking into account only the rows and columns with the numbers $q, q+1, ...r$. With our definition of the matrix $A(\lambda)$, its first row and column have the number equal to 0 while the last ones have the number $K = N - k$. Therefore the characteristic equation can be written as

$$\Delta_0^K(\lambda) = 0 \tag{6.11}$$

In general, since the field F_{p^2} is not algebraically closed, there is no guaranty that we will succeed in finding even one eigenvalue. However, we will see below that in a special case of the operator with the matrix elements (6.10), it is possible to find all $K + 1$ eigenvalues.

The matrix $A(\lambda)$ is three-diagonal. It is easy to see that

$$\Delta_0^{q+1}(\lambda) = -\lambda\Delta_0^q(\lambda) - A_{q,q+1}A_{q+1,q}\Delta_0^{q-1}(\lambda) \tag{6.12}$$

Let λ_l be a solution of Eq. (6.11). We denote $e_q \equiv e_{q+k,kl}$. Then the element

$$\chi(\lambda_l) = \sum_{q=0}^{K}\{(-1)^q\Delta_0^{q-1}(\lambda_l)e_q/[\prod_{s=0}^{q-1} A_{s,s+1}]\} \tag{6.13}$$

is the eigenvector of the operator $\mathcal{E}$ with the eigenvalue λ_l. This can be verified directly by using Eqs. (4.13) and (6.10-6.13).

To solve Eq. (6.12) we have to find the expressions for $\Delta_0^q(\lambda)$ when $q = 0, 1, ...K$. It is obvious that $\Delta_0^0(\lambda) = -\lambda$, and as follows from Eqs. (6.10) and (6.12),

$$\Delta_0^1(\lambda) = \lambda^2 - \frac{w + (2k+3)^2}{2(k+2)} \tag{6.14}$$

If $w = -\tilde{\mu}^2$ then it can be shown that $\Delta_0^q(\lambda)$ is given by the following expressions. If q is odd then

$$\Delta_0^q(\lambda) = \sum_{l=0}^{(q+1)/2} C^l_{(q+1)/2} \prod_{s=1}^{l} [\lambda^2 + (\tilde{\mu} - 2k - 4s + 1)^2](-1)^{(q+1)/2-l}$$
$$\prod_{s=l+1}^{(q+1)/2} \frac{(2k+2s+1)(\tilde{\mu} - 2k - 4s + 1)(\tilde{\mu} - 2k - 4s - 1)}{2(k + (q+1)/2 + s)} \tag{6.15}$$

and if q is even then

$$\Delta_0^q(\lambda) = (-\lambda) \sum_{l=0}^{q/2} C^l_{q/2} \prod_{s=1}^{l} [\lambda^2 + (\tilde{\mu} - 2k - 4s + 1)^2](-1)^{q/2-l}$$
$$\prod_{s=l+1}^{(q+1)/2} \frac{(2k+2s+1)(\tilde{\mu} - 2k - 4s - 1)(\tilde{\mu} - 2k - 4s - 3)}{2(k + q/2 + s + 1)} \tag{6.16}$$

Indeed, for $q = 0$ Eq. (6.16) is compatible with $\Delta_0^0(\lambda) = -\lambda$, and for $q = 1$ Eq. (6.15) is compatible with Eq. (6.14). Then one can directly verify that Eqs. (6.15) and (6.16) are compatible with Eq. (6.12).

With our definition of $\tilde{\mu}$, the only possibility for K is such that

$$\tilde{\mu} = 2K + 2k + 3 \tag{6.17}$$

Then, as follows from Eqs. (6.15) and (6.16), when K is odd or even, only the term with $l = [(K+1)/2]$ (where $[(K+1)/2]$ is the integer part of $(K+1)/2$) contributes to $\Delta_0^K(\lambda)$ and, as a consequence

$$\Delta_0^K(\lambda) = (-\lambda)^{r(K)} \prod_{k=1}^{[(K+1)/2]} [\lambda^2 + (\tilde{\mu} - 2j - 4k + 1)^2] \tag{6.18}$$

where $r(K) = 0$ if K is odd and $r(K) = 1$ if K is even. If $p = 3$ $(mod\ 4)$, this equation has solutions only if F_p is extended, and the minimum extension is F_{p^2}. Then the solutions are given by

$$\lambda = \pm i(\tilde{\mu} - 2k - 4s + 1) \quad (s = 1, 2...[(K+1)/2]) \tag{6.19}$$

and when K is even there also exists an additional solution $\lambda = 0$. When K is odd, solutions can be represented as

$$\lambda = \pm 2i,\ \pm 6i, ... \pm 2iK \tag{6.20}$$

while when K is even, the solutions can be represented as

$$\lambda = 0,\ \pm 4i,\ \pm 8i, ... \pm 2iK \tag{6.21}$$

Therefore the spectrum is equidistant and the distance between the neighboring elements is equal to $4i$. As follows from Eqs. (6.17), all the roots are simple and then, as follows from Eq. (6.13), the operator $\mathcal{E}$ is fully decomposable. It can be shown by a direct calculation [42] that the eigenvectors e corresponding to pure imaginary eigenvalues are such that $(e, e) = 0$ in F_p. Such a possibility has been mentioned in the preceding section.

Our conclusion is that if $p = 3\,(mod\,4)$ then all the operators $(\mathcal{E}, \mathbf{N}, \mathbf{B}, \mathbf{J})$ are fully decomposable if F_p is extended to F_{p^2} but no further extension is necessary. This might be an argument explaining why standard theory is based on complex numbers. On the other hand, our conclusion is obtained by considering states where n is not necessarily small in comparison with $p^{1/2}$ and standard physical intuition does not work in this case. One might think that the solutions (6.20) and (6.21) for the eigenvalues of the dS Hamiltonian indicate that GFQT is unphysical since the Hamiltonian cannot have imaginary eigenvalues. However, such a conclusion is premature since in standard quantum theory the Hamiltonian of a free particle does not have normalized eigenstates (since the spectrum is pure continuous) and therefore for any realistic state the width of the energy distribution cannot be zero.

If A is an operator of a physical quantity in standard theory then the distribution of this quantity in some state can be calculated in two ways. First, one can find eigenvectors of A, decompose the state over those eigenvectors and then the coefficients of the decomposition describe the distribution. Another possibility is to calculate all moments of A, i.e. the mean value, the mean square deviation etc. Note that the moments do not depend on the choice of basis since they are fully defined by the action of the operator on the given state. A standard result of the probability theory (see e.g. Ref. [108]) is that the set of moments uniquely defines the moment distribution function, which in turn uniquely defines the distribution. However in practice there is no need to know all the moments since the number of experimental data is finite and knowing only several first moments is typically quite sufficient.

In GFQT the first method does not necessarily defines the distribution. In particular, the above results for the dS Hamiltonian show that its eigenvectors $\sum_{nkl} c(n,k,l)e_{nkl}$ are such that $c(n,k,l) \neq 0$ for all $n = k, ...N$, where N is at least of the order of $p^{1/2}$. Since the $c(n,k,l)$ are elements of F_{p^2}, their formal modulus cannot be less than 1 and therefore the formal norm of such eigenvectors cannot be much less than p (the equality $(e, e) = 0$ takes place since the scalar product is calculated in F_p). Therefore eigenvectors of the dS Hamiltonian do not have a probabilistic interpretation. On the other hand, as already noted, we can consider states $\sum_{nkl} c(n,k,l)e_{nkl}$ such that $c(n,k,l) \neq 0$ only if $n_{min} \leq n \leq n_{max}$ where $n_{max} \ll N$. Then the probabilistic interpretation for such states might be a good approximation if at least several first moments give reasonable physical results (see the discussion of probabilities in Sect. 6.1). In Chaps. 4 and 5 we discussed semiclassical approximation taking into account only the first two moments: the mean value and mean square deviation.

Chapter 7

Semiclassical states in modular representations

7.1 Semiclassical states in GFQT

For constructing semiclassical states in GFQT one should use the basis defined by Eq. (4.8) and the coefficients $c(n, k, \mu)$ should be elements of F_{p^2}. Such states should satisfy several criteria. First, as noted in the preceding chapter, the probabilistic interpretation can be valid only if the quantities $\rho_0(n, k, \mu) = (e_{nk\mu}, e_{nk\mu})$ defined by Eq. (4.12) are such that $f(\rho_0(n, k, \mu)) \geq 0$ and $f(\rho_0(n, k, \mu)) \ll p$ where f is the map from F_p to Z defined in Sec. 6.1.

By using the fact that spaces in quantum theory are projective one can replace the basis elements $e_{nk\mu}$ by $Ce_{nk\mu}$ where $C \in F_{p^2}$ is any nonzero constant. Then the matrix elements of the operators in the new basis are the same and the normalizations are defined by the quantities $\rho(n, k, \mu) = C\bar{C}\rho_0(n, k, \mu)$. As noted in the preceding chapter, this reflects the fact that only ratios of probabilities have a physical meaning. Hence for ensuring probabilistic interpretation one could try to find C such that the quantities $f(\rho(n, k, \mu))$ have the least possible values.

As follows from Eq. (4.12),

$$\rho_0(n, k, \mu) = (2k+1)! C_{2k}^{k-\mu} C_n^k C_{n+k+1}^k \prod_{j=1}^{n} [w + (2j+1)^2] \tag{7.1}$$

As noted in Chap. 6, a probabilistic interpretation can be possible only if $c(n, k, \mu) \neq 0$ for $n \in [n_{min}, n_{max}]$, $k \in [k_{min}, k_{max}]$ and $\mu \in [\mu_{min}, \mu_{max}]$. Hence our nearest goal is to find the constant C such that the quantities $\rho(n, k, \mu)$ have the least possible values when the quantum numbers $(nk\mu)$ are in the above range.

We denote $\Delta n = n_{max} - n_{min}$, $\Delta k = k_{max} - k_{min}$ and $\Delta\mu = \mu_{max} - \mu_{min}$. Since R is very large, we expect that $\Delta n \gg \Delta k, \Delta\mu$ but since the exact value of R is not known, we don't know whether a typical value of k is much greater than Δn or not. One can directly verify that $\rho_0(n,k,\mu) = C_1\rho(n,k,\mu)$ where

$$\rho(n,k,\mu) = 4^{k-k_{min}}\frac{(2k+1)!!(2k-1)!!(k_{max}-\mu_{min})!(k_{max}+\mu_{max})!}{(2k_{min}+1)!!(2k_{min}-1)!!(k-\mu)!(k+\mu)!}$$
$$\prod_{j=0}^{n+k-n_{min}-k_{min}-1}(n_{min}+k_{min}+2+j)\prod_{j=0}^{n_{max}-k_{min}-n+k-1}(n-k+1+j)$$
$$\prod_{j=0}^{n-n_{min}-1}(n_{min}+1+j)\prod_{j=0}^{n_{max}-n-1}(n+2+j)\prod_{j=n_{min}+1}^{n}[w+(2j+1)^2] \quad (7.2)$$

$$C_1 = 4^{k_{min}}\frac{(2k_{min}+1)!!(2k_{min}-1)!!}{(k_{max}-\mu_{min})!(k_{max}+\mu_{max})!}\prod_{j=1}^{n_{min}}[w+(2j+1)^2]$$
$$\prod_{j=0}^{k_{min}-\Delta n-1}(n_{max}+2+j)(n_{max}-k_{min}+1+j) \quad (7.3)$$

if $k \gg \Delta n$ and

$$\rho(n,k,\mu) = 4^{k-k_{min}}\frac{(2k+1)!!(2k-1)!!(k_{max}-\mu_{min})!(k_{max}+\mu_{max})!}{(2k_{min}+1)!!(2k_{min}-1)!!(k-\mu)!(k+\mu)!}$$
$$\prod_{j=0}^{k-1}[(n+2+j)(n-k+1+j)]\prod_{j=n_{min}+1}^{n}[w+(2j+1)^2] \quad (7.4)$$

$$C_1 = 4^{k_{min}}\frac{(2k_{min}+1)!!(2k_{min}-1)!!}{(k_{max}-\mu_{min})!(k_{max}+\mu_{max})!}\prod_{j=1}^{n_{min}}[w+(2j+1)^2] \quad (7.5)$$

if k is of the same order than Δn or less.

We now have to prove the existence of the constant C such that $C\bar{C} = C_2$ where $C_2 = 1/C_1$. For this purpose we note the following. It is known [98] that any Galois field without its zero element is a cyclic multiplicative group. Let r be a primitive root in F_p, *i.e.*, the element such that any nonzero element of F_p can be represented as r^s $(s = 1,2,...,p-1)$. Hence, if $C_2 = r^s$ and s is even then $C = r^{s/2}$ obviously satisfies the above requirement.

Suppose now that s is odd. As noted in Chap. 6, -1 is a quadratic residue in F_p if $p = 1$ $(mod\ 4)$ and a quadratic non-residue in F_p if $p = 3$ $(mod\ 4)$. Therefore in the case $p = 3$ $(mod\ 4)$ we have $-1 = r^q$ where q is odd. Hence $C_2 = -C_3$ where $C_3 = r^{s+q}$ is a quadratic residue in F_p. Now the quantity C satisfying the above requirement exists if $C = \alpha r^{(s+q)2}$ and α satisfies the equation

$$\alpha\bar{\alpha} = -1 \quad (7.6)$$

For proving that the solution of this equation exists we again use the property that any Galois field without its zero element is a cyclic multiplicative group but now this property is applied in the case of F_{p^2} with $p = 3\ (mod\ 4)$. Let now r be a primitive root in F_{p^2}. It is known [98] that the only nontrivial automorphism of F_{p^2} is $\alpha \to \bar{\alpha} = \alpha^p$. Therefore if $\alpha = r^s$ then $\alpha\bar{\alpha} = r^{(p+1)s}$. On the other hand, since $r^{(p^2-1)} = 1$, $r^{(p^2-1)/2} = -1$. Therefore a solution of Eq. (7.6) exists at least with $s = (p-1)/2$.

The next step is to investigate conditions for the coefficients $c(n,k,\mu)$ such that the state $\sum_{nk\mu} c(n,k,\mu)e(n,k,\mu)$ is semiclassical. As noted in Sect. 4.2, in standard theory the quantities $c(n,k,\mu)$ contain the factor $exp[i(-n\varphi + k\alpha - \mu\beta)]$ and in the region of maximum the quantities $|c(n,k,\mu)|^2$ are of the same order. To generalize these conditions to the case of GFQT we define a function F from the set of complex numbers to F_{p^2}. If a is a real number then we define $l = Round(a)$ as an integer closest to a. This definition is ambiguous when $a = l \pm 0.5$ but in the region of maximum the numbers in question are very large and the rounding errors ± 1 are not important. Analogously, if $z = a + bi$ is a complex number then we define $Round(z) = Round(a) + Round(b)i$. Finally, we define $F(z) \in F_{p^2}$ as $f(Round(z))$.

As follows from Eqs. (7.2) and (7.4), the quantity $\rho(n,k,\mu)$ has the maximum at $n = n_{max}$, $k = k_{max}$, $\mu = \mu_{min}$. Consider the state $\sum_{nk\mu} c(n,k,\mu)e(n,k,\mu)$ such that

$$c(n,k,\mu) = a(n,k,\mu)F\{[\frac{\rho(n_{max},k_{max},\mu_{min}}{\rho(n,k,\mu)}]^{1/2}exp[i(-n\varphi + k\alpha - \mu\beta)]\} \qquad (7.7)$$

where $a(n,k,\mu)$ is a slowly changing function in the region of maximum.

For the validity of semiclassical approximation the condition

$$\rho(n_{max},k_{max},\mu_{min})\sum_{nk\mu}|a(n,k,\mu)|^2 \ll p \qquad (7.8)$$

should be satisfied. As follows from Eqs. (7.2) and (7.4), for a nonrelativistic particle it will be satisfied if

$$(4k_{max})^{\Delta k}[(k_{max}-\mu_{min})(k_{max}+\mu_{max})]^{(\Delta k+\Delta\mu)}n_{max}^{2(\Delta n+\Delta k)}w^{\Delta n}A\Delta n\Delta k\Delta\mu \ll p \qquad (7.9)$$

or

$$(4k_{max})^{\Delta k}[(k_{max}-\mu_{min})(k_{max}+\mu_{max})]^{(\Delta k+\Delta\mu)}w^{\Delta n}A\Delta n\Delta k\Delta\mu \ll p \qquad (7.10)$$

respectively, where A is the maximum value of $|a(n,k,\mu)|^2$. If A is not anomalously large then in the both cases those conditions can be approximately written as

$$\Delta n ln w \ll ln p \qquad (7.11)$$

Therefore not only the number p should be very large, but even lnp should be very large.

7.2 Many-body systems in GFQT and gravitational constant

In quantum theory, state vectors of a system of N bodies belong to the Hilbert space which is the tensor product of single-body Hilbert spaces. This means that state vectors of the N-body systems are all possible linear combinations of functions

$$\psi(n_1,k_1,l_1,...n_N,k_N,l_N)=\psi_1(n_1,k_1,l_1)\cdots\psi_N(n_N,k_N,l_N) \tag{7.12}$$

By definition, the bodies do not interact if all representation operators of the symmetry algebra for the N-body systems are sums of the corresponding single-body operators. For example, the energy operator $\mathcal{E}$ for the N-body system is a sum $\mathcal{E}_1+\mathcal{E}_2+...+\mathcal{E}_N$ where the operator $\mathcal{E}_i$ $(i=1,2,...N)$ acts nontrivially over its "own" variables (n_i,k_i,l_i) while over other variables it acts as the identity operator.

If we have a system of noninteracting bodies in standard quantum theory, each $\psi_i(n_i,k_i,l_i)$ in Eq. (7.12) is fully independent of states of other bodies. However, in GFQT the situation is different. Here, as shown in the preceding section, a necessary condition for a wave function to have a probabilistic interpretation is given by Eq. (7.11). Since we assume that p is very large, this is not a serious restriction. However, if a system consists of N components, a necessary condition that the wave function of the system has a probabilistic interpretation is

$$\sum_{i=1}^{N}\delta_i lnw_i \ll lnp \tag{7.13}$$

where $\delta_i=\Delta n_i$ and $w_i=4R^2m_i^2$ where m_i is the mass of the subsystem i. This condition shows that in GFQT the greater the number of components is, the stronger is the restriction on the width of the dS momentum distribution for each component. This is a crucial difference between standard theory and GFQT. A naive explanation is that if p is finite, the same set of numbers which was used for describing one body is now shared between N bodies. In other words, if in standard theory each body in the free N-body system does not feel the presence of other bodies, in GFQT this is not the case. This might be treated as an effective interaction in the free N-body system.

In Chaps. 3 and 5 we discussed a system of two free bodies such their relative motion can be described in the framework of semiclassical approximation. We have shown that the mean value of the mass operator for this system differs from the expression given by standard Poincare theory. The difference describes an effective interaction which we treat as the dS antigravity at very large distances and gravity when the distances are much less than cosmological ones. In the latter case the result depends on the total dS momentum distribution for each body (see Eq. (5.31)). Since the interaction is proportional to the masses of the bodies, this effect is important only in situations when at least one body is macroscopic. Indeed, if neither of the bodies is macroscopic, their masses are small and their relative motion is not described in the framework of semiclassical approximation. In particular, in this approach, gravity between two elementary particles has no physical meaning.

The existing quantum theory does not make it possible to reliably calculate the width of the total dS momentum distribution for a macroscopic body and at best only a qualitative estimation of this quantity can be given. The above discussion shows that the greater is the mass of the macroscopic body, the stronger is the restriction on the dS momentum distribution for each subsystem of this body. Suppose that a body with the mass M can be treated as a composite system consisting of similar subsystems with the mass m. Then the number of subsystems is $N = M/m$ and, as follows from Eq. (7.13), the width δ of their dS momentum distributions should satisfy the condition $N\delta lnw \ll lnp$ where $w = 4R^2m^2$. Since the greater the value of δ is, the more accurate is the semiclassical approximation, a reasonable scenario is that each subsystem tends to have the maximum possible δ but the above restriction allows to have only such value of δ that it is of the order of magnitude not exceeding $lnp/(Nlnw)$.

The next question is how to estimate the width of the total dS momentum distribution for a macroscopic body. For solving this problem one has to change variables from individual dS momenta of subsystems to total and relative dS momenta. Now the total dS momentum and relative dS momenta will have their own momentum distributions which are subject to a restriction similar to that given by Eq. (7.13). If we assume that all the variables share this restriction equally then the width of the total momentum distribution also will be a quantity not exceeding $lnp/(Nlnw)$. Suppose that $m = N_1m_0$ where m_0 is the nucleon mass. The value of N_1 should be such that our subsystem still can be described by semiclassical approximation. Then the estimation of δ is

$$\delta = N_1m_0lnp/[2Mln(2RN_1m_0)] \tag{7.14}$$

Suppose that N_1 can be taken to be the same for all macroscopic bodies. For example, it is reasonable to expect that when N_1 is of the order of 10^3, the subsystems still can be described by semiclassical approximation but probably this is the case even for smaller values of N_1.

In summary, although calculation of the width of the total dS momentum distribution for a macroscopic body is a very difficult problem, GFQT gives a reasonable qualitative explanation why this quantity is inversely proportional to the mass of the body. With the estimation (7.14), the result given by Eq. (5.31) can be written in the form (5.33) where

$$G = \frac{2const\,Rln(2RN_1m_0)}{N_1m_0lnp} \tag{7.15}$$

In Chaps. 1 and 6 we argued that in theories based on dS invariance and/or Galois fields, neither the gravitational nor cosmological constant can be fundamental. In particular, in units $\hbar/2 = c = 1$, the dimension of G is $length^2$ and its numerical value is l_P^2 where l_P is the Planck length ($l_P \approx 10^{-35}m$). Equation (7.15) is an additional indication that this is the case since G depends on R (or the cosmological constant) and there is no reason to think that it does not change with time. Since $G_{dS} = G\Lambda$ is dimensionless in units $\hbar/2 = c = 1$, this quantity should be treated as the gravitational constant in dS theory. Let $\mu = 2Rm_0$ be the dS nucleon mass and $\Lambda = 3/R^2$ be the cosmological constant. Then Eq. (7.15) can be written as

$$G_{dS} = \frac{12const\, ln(N_1\mu)}{N_1\mu lnp} \qquad (7.16)$$

As noted in Sect. 1.4, standard cosmological constant problem arises when one tries to explain the value of Λ from quantum theory of gravity assuming that this theory is QFT, G is fundamental and dS symmetry is a manifestation of dark energy (or other fields) on flat Minkowski background. Such a theory contains strong divergences and the result depends on the value of the cutoff momentum. With a reasonable assumption about this value, the quantity Λ is of the order of $1/G$ and this is reasonable since G is the only parameter in this theory. Then Λ is by more than 120 orders of magnitude greater than its experimental value. However, in our approach we have an additional fundamental parameter p. Equation (7.16) shows that $G\Lambda$ is not of the order of unity but is very small since not only p but even lnp is very large. For a rough estimation, we assume that the values of $const$ and N_1 in this expression are of the order of unity. Then if, for example, R is of the order of $10^{26}m$, we have that μ is of the order of 10^{42} and lnp is of the order of 10^{80}. Therefore p is a huge number of the order of $exp(10^{80})$. In the preceding chapter we argued that standard theory can be treated as a special case of GFQT in the formal limit $p \to \infty$. The above discussion shows that restrictions on the width of the total dS momentum arise because p is not infinitely large. It is seen from Eq. (7.16) that gravity disappears in the above formal limit. Therefore in our approach gravity is a consequence of the fact that dS symmetry is considered over a Galois field rather than the field of complex numbers.

Chapter 8

Basic properties of AdS quantum theories

As noted in Sec. 3.1, if one considers Poincare, AdS and dS symmetries in standard theory then only the latter symmetry does not contradict the possibility that gravity can be described in the framework of a free theory. In addition, as shown in Secs. 3.6 and 5.1, the fact that $\Lambda > 0$ can be treated simply as an indication that among the three symmetries the dS one is the most pertinent for describing nature.

In standard theory the difference between IRs of the so(2,3) and so(1,4) algebras is that an IR of the so(2,3) algebra where the operators $M^{\mu 4}$ ($\mu = 0, 1, 2, 3$) are Hermitian can be treated as IRs of the so(1,4) algebra where these operators are anti-Hermitian and vice versa. As noted in Chap. 6, in GFQT a probabilistic interpretation is only approximate and hence Hermiticy can be only a good approximation in some situations. Therefore one cannot exclude a possibility that elementary particles can be described by modular analogs of IRs of the so(2,3) algebra while modular representations describing symmetry of macroscopic bodies are modular analogs of standard representations of the so(1,4) algebra.

In this chapter standard and modular IRs of the so(2,3) algebra are discussed in parallel in order to demonstrate common features and differences between standard and modular cases.

8.1 Modular IRs of the sp(2) and su(2) algebra

The key role in constructing modular IRs of the so(2,3) algebra is played by modular IRs of the sp(2) subalgebra. They are described by a set of operators $(a', a", h)$ satisfying the commutation relations

$$[h, a'] = -2a', \quad [h, a"] = 2a", \quad [a', a"] = h \tag{8.1}$$

The Casimir operator of the second order for the algebra (8.1) has the form

$$K = h^2 - 2h - 4a"a' = h^2 + 2h - 4a'a" \tag{8.2}$$

We first consider representations with the vector e_0 such that

$$a'e_0 = 0, \quad he_0 = q_0 e_0 \tag{8.3}$$

where $q_0 \in F_p$. We will denote q_0 by the numbers $0, 1, ...p-1$. In general we consider the representation in a linear space over F_{p^k} where k is a natural number (see Sec. 6.1). Denote $e_n = (a")^n e_0$. Then it follows from Eqs. (8.2) and (8.3), that

$$he_n = (q_0 + 2n)e_n, \quad Ke_n = q_0(q_0 - 2)e_n \tag{8.4}$$

$$a'a"e_n = (n+1)(q_0 + n)e_n \tag{8.5}$$

One can consider analogous representations in standard theory. Then q_0 is a positive real number, $n = 0, 1, 2, ...$ and the elements e_n form a basis of the IR. In this case e_0 is a vector with a minimum eigenvalue of the operator h (minimum weight) and there are no vectors with the maximum weight. The operator h is positive definite and bounded below by the quantity q_0. For these reasons the above modular IRs can be treated as modular analogs of such standard IRs that h is positive definite.

Analogously, one can construct modular IRs starting from the element e_0' such that

$$a"e_0' = 0, \quad he_0' = -q_0 e_0' \tag{8.6}$$

and the elements e_n' can be defined as $e_n' = (a')^n e_0'$. Such modular IRs are analogs of standard IRs where h is negative definite. However, in the modular case Eqs. (8.3) and (8.6) define the same IRs. This is clear from the following consideration.

The set $(e_0, e_1, ...e_N)$ will be a basis of IR if $a"e_i \neq 0$ for $i < N$ and $a"e_N = 0$. These conditions must be compatible with $a'a"e_N = 0$. The case $q_0 = 0$ is of no interest since, as follows from Eqs. (8.3-8.6), all the representation operators are null operators, the representation is one-dimensional and e_0 is the only basis vector in the representation space. If $q_0 = 1, ...p-1$, it follows from Eq. (8.5) that N is defined by the condition $q_0 + N = 0$. Hence $N = p - q_0$ and the dimension of IR equals

$$Dim(q_0) = p - q_0 + 1 \tag{8.7}$$

This result is formally valid for all the values of q_0 if we treat q_0 as one of the numbers $1, ...p-1, p$. It is easy to see that e_N satisfies Eq. (8.6) and therefore it can be identified with e_0'.

Let us forget for a moment that the eigenvalues of the operator h belong to F_p and will treat them as integers. Then, as follows from Eq. (8.4), the eigenvalues are

$$q_0, q_0 + 2, ..., 2p - 2 - q_0, 2p - q_0.$$

Therefore, if $f(q_0) > 0$ and $f(q_0) \ll p$, the maximum value of q_0 is $2p - q_0$, i.e. it is of the order of $2p$.

In standard theory, IRs are discussed in Hilbert spaces, i.e. the space of the IR is supplied by a positive definite scalar product. It can be defined such that

$(e_0, e_0) = 1$, the operator h is self-adjoint and the operators a' and $a"$ are adjoint to each other: $(a')^* = a"$. Then, as follows from Eq. (8.5),

$$(e_n, e_n) = n!(q_0)_n \tag{8.8}$$

where we use the Pocchammer symbol $(q_0)_n = q_0(q_0 + 1) \cdots (q_0 + n - 1)$. Usually the basis vectors are normalized to one but this is only a matter of convention but not a matter of principle since not the probability itself but only ratios of probabilities have a physical meaning (see the discussion in Chap. 6). In GFQT one can formally define the scalar product by the same formulas but in that case this scalar product cannot be positive definite since in Galois fields the notions of positive and negative numbers can be only approximate. Therefore, as noted in Chap. 6 in GFQT the probabilistic interpretation cannot be universal. However, if the quantities q_0 and n are such that the r.h.s. of Eq. (8.8) is much less than p then the probabilistic interpretation is (approximately) valid if the IR is discussed in a space over F_{p^2} (see Chap. 6 for a detailed discussion). Therefore if p is very large, then for a large number of elements there is a correspondence between standard theory and GFQT

Representations of the su(2) algebra are defined by a set of operators (L_+, L_-, L_3) satisfying the commutations relations

$$[L_3, L_+] = 2L_+, \quad [L_3, L_-] = 2L_-, \quad [L_+, L_-] = 2L_3 \tag{8.9}$$

In the case of representations over the field of complex numbers, these relations can be formally obtained from Eq. (8.1) by the replacements $h \to L_3$, $a' \to iL_-$ and $a" \to iL_+$. The difference between the representations of the sp(2) and su(2) algebras in Hilbert spaces is that in the latter case the Hermiticity conditions are $L_3^* = L_3$ and $L_+^* = L_-$. The Casimir operator for the algebra (8.9) is

$$K = L_3^2 - 2L_3 + 4L_+L_- = L_3^2 + 2L_3 + 4L_-L_+ \tag{8.10}$$

For constructing IRs, we assume that the representation space contains a vector e_0 such that

$$L_3 e_0 = s e_0 \quad L_+ e_0 = 0 \tag{8.11}$$

where $s \geq 0$ for standard IRs and $s \in F_p$ for modular IRs. In the letter case we will denote s by the numbers $0, 1, ...p - 1$. If $e_k = (L_-)^k e_0$ $(k = 0, 1, 2, ...)$ then it is easy to see that

$$L_3 e_k = (s - 2k)e_k, \quad K e_k = s(s + 2)e_k, \quad L_+L_-e_k = (k + 1)(s - k)e_k \tag{8.12}$$

The IR will be finite dimensional if there exists $k = k_{max}$ such that $L_+L_-e_k = 0$ for this value of k. As follows from the above expression, for modular IRs such a value of k always exists, $k_{max} = s$ and the dimension of the IR is $Dim(s) = s + 1$. For standard IRs the same conclusion is valid if s iz zero or a natural number. In standard quantum theory, the representation operators of the su(2) algebra are associated with the components of the angular momentum operator $\mathbf{L} = (L_x, L_y, L_z)$ such that $L_3 = L_z$ and $L_\pm = (L_x \pm iL_y)/2$. The commutation relations for the components of $\mathbf{L}$ are usually written in units where $\hbar = 1$. Then s can be only an integer or a half-integer and $Dim(s) = 2s + 1$.

8.2 Modular IRs of the so(2,3) Algebra

Standard IRs of the so(2,3) algebra relevant for describing elementary particles have been considered by several authors. The description in this section is a combination of two elegant ones given in Ref. [80] for standard IRs and Ref. [84] for modular IRs. As already noted, in standard theory, the commutation relations between the representation operators are given by Eq. (4.1) where $\eta^{44} = \pm 1$ for the AdS and dS cases, respectively. As follows from the contraction procedure described in Sec. 1.3, the operator M^{04} can be treated as the AdS analog of the energy operator.

If a modular IR is considered in a linear space over F_{p^2} with $p = 3\ (mod\ 4)$ then Eq. (4.1) is also valid. However, in the general case we consider modular IRs in linear spaces over F_{p^k} where k is arbitrary. In this case it is convenient to work with another set of ten operators. Let $(a'_j, a_j", h_j)$ $(j = 1, 2)$ be two independent sets of operators satisfying the commutation relations for the sp(2) algebra

$$[h_j, a'_j] = -2a'_j, \quad [h_j, a_j"] = 2a_j", \quad [a'_j, a_j"] = h_j \tag{8.13}$$

The sets are independent in the sense that for different j they mutually commute with each other. We denote additional four operators as $b', b", L_+, L_-$. The operators $L_3 = h_1 - h_2, L_+, L_-$ satisfy the commutation relations (8.9) of the su(2) algebra while the other commutation relations are as follows

$$\begin{gathered}
[a'_1, b'] = [a'_2, b'] = [a_1", b"] = [a_2", b"] = [a'_1, L_-] = [a_1", L_+] = \\
[a'_2, L_+] = [a_2", L_-] = 0, \quad [h_j, b'] = -b', \quad [h_j, b"] = b" \\
[h_1, L_\pm] = \pm L_\pm, \quad [h_2, L_\pm] = \mp L_\pm, \quad [b', b"] = h_1 + h_2 \\
[b', L_-] = 2a'_1, \quad [b', L_+] = 2a'_2, \quad [b", L_-] = -2a_2", \quad [b", L_+] = -2a_1" \\
[a'_1, b"] = [b', a_2"] = L_-, \quad [a'_2, b"] = [b', a_1"] = L_+ \\
[a'_1, L_+] = [a'_2, L_-] = b', \quad [a_2", L_+] = [a_1", L_-] = -b"
\end{gathered} \tag{8.14}$$

At first glance these relations might seem rather chaotic but in fact they are very natural in the Weyl basis of the so(2,3) algebra.

In spaces over F_{p^2} with $p = 3\ (mod\ 4)$ the relation between the above sets of ten operators is

$$\begin{gathered}
M_{10} = i(a_1" - a'_1 - a_2" + a'_2), \quad M_{14} = a_2" + a'_2 - a_1" - a'_1 \\
M_{20} = a_1" + a_2" + a'_1 + a'_2, \quad M_{24} = i(a_1" + a_2" - a'_1 - a'_2) \\
M_{12} = L_3, \quad M_{23} = L_+ + L_-, \quad M_{31} = -i(L_+ - L_-) \\
M_{04} = h_1 + h_2, \quad M_{34} = b' + b", \quad M_{30} = -i(b" - b')
\end{gathered} \tag{8.15}$$

and therefore the sets are equivalent. However, the relations (8.9,8.13,8.14) are more general since they can be used when the representation space is a space over F_{p^k} with an arbitrary k. It is also obvious that such a *definition* of the operators M_{ab} is not unique. For example, any cyclic permutation of the indices $(1,2,3)$ gives a new set of operators satisfying the same commutation relations.

In standard theory, the Casimir operator of the second order for the representation of the so(2,3) algebra is given by

$$I_2 = \frac{1}{2}\sum_{ab} M_{ab}M^{ab} \tag{8.16}$$

As follows from Eqs. (8.9,8.13-8.15), I_2 can be written as

$$I_2 = 2(h_1^2 + h_2^2 - 2h_1 - 4h_2 - 2b"b' + 2L_-L_+ - 4a_1"a_1' - 4a_2"a_2') \tag{8.17}$$

We use the basis in which the operators (h_j, K_j) $(j = 1,2)$ are diagonal. Here K_j is the Casimir operator (8.2) for the algebra $(a_j', a_j", h_j)$. For constructing IRs we need operators relating different representations of the sp(2)×sp(2) algebra. By analogy with Refs. [80, 84], one of the possible choices is as follows

$$\begin{aligned} A^{++} &= b"(h_1-1)(h_2-1) - a_1"L_-(h_2-1) - a_2"L_+(h_1-1) + a_1"a_2"b' \\ A^{+-} &= L_+(h_1-1) - a_1"b', \quad A^{-+} = L_-(h_2-1) - a_2"b', \quad A^{--} = b' \end{aligned} \tag{8.18}$$

We consider the action of these operators only on the space of minimal sp(2)×sp(2) vectors, *i.e.* such vectors x that $a_j'x = 0$ for $j = 1,2$, and x is the eigenvector of the operators h_j. If x is a minimal vector such that $h_jx = \alpha_jx$ then $A^{++}x$ is the minimal eigenvector of the operators h_j with the eigenvalues $\alpha_j + 1$, $A^{+-}x$ - with the eigenvalues $(\alpha_1 + 1, \alpha_2 - 1)$, $A^{-+}x$ - with the eigenvalues $(\alpha_1 - 1, \alpha_2 + 1)$, and $A^{--}x$ - with the eigenvalues $\alpha_j - 1$.

By analogy with Refs. [80, 84], we require the existence of the vector e_0 satisfying the conditions

$$a_j'e_0 = b'e_0 = L_+e_0 = 0, \quad h_je_0 = q_je_0 \quad (j = 1,2) \tag{8.19}$$

where $q_j \in F_p$. As follows from Eq. (8.17), in the IR characterized by the quantities (q_1, q_2), all the nonzero elements of the representation space are the eigenvectors of the operator I_2 with the eigenvalue

$$I_2 = 2(q_1^2 + q_2^2 - 2q_1 - 4q_2) \tag{8.20}$$

Since $L_3 = h_1 - h_2$ then, as follows from the results of Sec. 8.1, if q_1 and q_2 are characterized by the numbers $0, 1, ...p-1$, $q_1 \geq q_2$ and $q_1 - q_2 = s$ then the elements $(L_+)^ke_0$ $(k = 0, 1, ...s)$ form a basis of the IR of the su(2) algebra with the spin s such that the dimension of the IR is $s + 1$. Therefore in the theory over a

Galois field the case when $q_1 < q_2$ should be treated such that $s = p + q_1 - q_2$. IRs with $q_1 < q_2$ have no analogs in standard theory and we will call them special IRs.

As follows from Eqs. (8.13) and (8.14), the operators (a_1', a_2', b') reduce the AdS energy $(h_1 + h_2)$ by two units. Therefore e_0 is an analog the state with the minimum energy which can be called the rest state. For this reason we use m_{AdS} to denote $q_1 + q_2$. In standard classification [80], the massive case is characterized by the condition $q_2 > 1$ and the massless one—by the condition $q_2 = 1$. Hence in standard theory the quantity m_{AdS} in the massive case is always greater than 2. There also exist two exceptional IRs discovered by Dirac [109] (Dirac singletons). They are characterized by the conditions $(m_{AdS} = 1, s = 0)$ and $(m_{AdS} = 2, s = 1)$ or in terms of (q_1, q_2), by the conditions $(q_1 = 1/2, q_2 = 1/2)$ and $(q_1 = 3/2, q_2 = 1/2)$, respectively.

In the theory over a Galois field 1/2 should be treated as $(p + 1)/2$ and 3/2 — as $(p + 3)/2$. Hence the Dirac singletons are characterized by the conditions $(q_1 = (p + 1)/2, q_2 = (p + 1)/2)$ and $(q_1 = (p + 3)/2, q_2 = (p + 1)/2)$, respectively. In general, in this theory it is possible that the quantities (q_1, q_2) are given by the numbers $2, 3, ...p - 1$ but since $q_1 + q_2$ is taken modulo p, it is possible that m_{AdS} can take one of the values (0,1,2). These cases also have no analogs in standard theory and we will call them special singleton IRs but will not treat the Dirac singletons as special. In this section we will consider the massive case while the singleton, massless and special cases will be considered in the next section.

As follows from the above remarks, the elements

$$e_{nk} = (A^{++})^n (A^{-+})^k e_0 \tag{8.21}$$

represent the minimal sp(2)×sp(2) vectors with the eigenvalues of the operators h_1 and h_2 equal to $Q_1(n, k) = q_1 + n - k$ and $Q_2(n, k) = q_2 + n + k$, respectively.

Consider the element $A^{--}A^{++}e_{nk}$. In view of the properties of the A operators mentioned above, this element is proportional to e_{nk} and therefore one can write $A^{--}A^{++}e_{nk} = a(n, k)e_{nk}$. One can directly verify that the actions of the operators A^{++} and A^{-+} on the space of minimal $sp(2)\times sp(2)$ vectors are commutative and therefore $a(n, k)$ does not depend on k. A direct calculation gives

$$(A^{--}A^{++} - A^{++}A^{--})e(n, k) = \{(Q_2 - 1)[Q_1 - 1)(Q_1 + Q_2) - (Q_1 - Q_2)] +$$
$$(Q_1 + Q_2 - 2)(\frac{1}{2}Q_1^2 + \frac{1}{2}Q_2^2 - Q_1 - 2Q_2 - \frac{1}{4}I_2)\}e(n, k) \tag{8.22}$$

where $Q_1 \equiv Q_1(n, k)$ and $Q_2 \equiv Q_2(n, k)$. As follows from this expression,

$$a(n) - a(n - 1) = q_1(q_2 - 1)(m_{AdS} - 2) + 2n(q_1^2 + q_2^2 +$$
$$3q_1q_2 - 5q_1 - 4q_2 + 4) + 6n^2(m_{AdS} - 2) + 4n^3 \tag{8.23}$$

Since $b'e_0 = 0$ by construction, we have that $a(-1) = 0$ and a direct calculation shows that, as a consequence of Eq. (8.23)

$$a(n) = (n + 1)(m_{AdS} + n - 2)(q_1 + n)(q_2 + n - 1) \tag{8.24}$$

Analogously, one can write $A^{+-}A^{-+}e_{nk} = b(k)e_{nk}$ and the result of a direct calculation is

$$b(k) = -\frac{1}{4}(Q_1 - 2)(Q_2 - 1)(2Q_1^2 + 2Q_2^2 - 8Q_1 - 4Q_2 - I_2) + a(n-1) \quad (8.25)$$

Then, as a consequence of Eqs. (8.20) and (8.24)

$$b(k) = (k+1)(s-k)(q_1 - k - 2)(q_2 + k - 1) \quad (8.26)$$

As follows from these expressions, in the massive case k can assume only the values $0, 1, ...s$ and in standard theory $n = 0, 1, ...\infty$. However, in the modular case $n = 0, 1, ...n_{max}$ where n_{max} is the first number for which the r.h.s. of Eq. (8.24) becomes zero in F_p. Therefore $n_{max} = p + 2 - m_{AdS}$.

The full basis of the representation space can be chosen in the form

$$e(n_1 n_2 nk) = (a_1")^{n_1}(a_2")^{n_2} e_{nk} \quad (8.27)$$

In standard theory n_1 and n_2 can be any natural numbers. However, as follows from the results of the preceding section, Eq. (8.13) and the properties of the A operators,

$$\begin{aligned} n_1 &= 0, 1, ...N_1(n,k), \quad n_2 = 0, 1, ...N_2(n,k) \\ N_1(n,k) &= p - q_1 - n + k, \quad N_2(n,k) = p - q_2 - n - k \end{aligned} \quad (8.28)$$

As a consequence, the representation is finite dimensional in agreement with the Zassenhaus theorem [105] (moreover, it is finite since any Galois field is finite).

Let us assume additionally that the representation space is supplied by a scalar product (see Chap. 6). The element e_0 can always be chosen such that $(e_0, e_0) = 1$. Suppose that the representation operators satisfy the Hermiticity conditions $L_+^* = L_-$, $a_j^{'*} = a_j"$, $b^{'*} = b"$ and $h_j^* = h_j$. Then, as follows from Eq. (8.15), in a special case when the representation space is a space over F_{p^2} with $p = 3$ $(mod\ 4)$, the operators M^{ab} are Hermitian as it should be. By using Eqs. (8.13-8.26), one can show by a direct calculation that the elements $e(n_1 n_2 nk)$ are mutually orthogonal while the quantity

$$Norm(n_1 n_2 nk) = (e(n_1 n_2 nk), e(n_1 n_2 nk)) \quad (8.29)$$

can be represented as

$$Norm(n_1 n_2 nk) = F(n_1 n_2 nk) G(nk) \quad (8.30)$$

where

$$\begin{aligned} F(n_1 n_2 nk) &= n_1!(Q_1(n,k) + n_1 - 1)!n_2!(Q_2(n,k) + n_2 - 1)! \\ G(nk) &= \{(q_2 + k - 2)!n!(m_{AdS} + n - 3)!(q_1 + n - 1)!(q_2 + n - 2)!k!s!\} \\ &\{(q_1 - k - 2)![(q_2 - 2)!]^3(q_1 - 1)!(m_{AdS} - 3)!(s-k)! \\ &[Q_1(n,k) - 1][Q_2(n,k) - 1]\}^{-1} \end{aligned} \quad (8.31)$$

In standard Poincare and AdS theories there also exist IRs with negative energies. They can be constructed by analogy with positive energy IRs. Instead of Eq. (8.19) one can require the existence of the vector e_0' such that

$$a_j"e_0' = b"e_0' = L_-e_0' = 0, \quad h_je_0' = -q_je_0', \quad (e_0', e_0') \neq 0 \quad (j = 1, 2) \qquad (8.32)$$

where the quantities q_1, q_2 are the same as for positive energy IRs. It is obvious that positive and negative energy IRs are fully independent since the spectrum of the operator M^{04} for such IRs is positive and negative, respectively. However, *the modular analog of a positive energy IR characterized by* q_1, q_2 *in Eq. (8.19), and the modular analog of a negative energy IR characterized by the same values of* q_1, q_2 *in Eq. (8.32) represent the same modular IR.* This is the crucial difference between standard quantum theory and GFQT, and a proof is given below.

Let e_0 be a vector satisfying Eq. (8.19). Denote $N_1 = p - q_1$ and $N_2 = p - q_2$. Our goal is to prove that the vector $x = (a_1")^{N_1}(a_2")^{N_2}e_0$ satisfies the conditions (8.32), *i.e.* x can be identified with e_0'.

As follows from the definition of N_1, N_2, the vector x is the eigenvector of the operators h_1 and h_2 with the eigenvalues $-q_1$ and $-q_2$, respectively, and, in addition, it satisfies the conditions $a_1"x = a_2"x = 0$. Let us prove that $b"x = 0$. Since $b"$ commutes with the $a_j"$, we can write $b"x$ in the form

$$b"x = (a_1")^{N_1}(a_2")^{N_2}b"e_0 \qquad (8.33)$$

As follows from Eqs. (8.14) and (8.19), $a_2'b"e_0 = L_+e_0 = 0$ and $b"e_0$ is the eigenvector of the operator h_2 with the eigenvalue $q_2 + 1$. Therefore, $b"e_0$ is the minimal vector of the sp(2) IR which has the dimension $p - q_2 = N_2$. Hence $(a_2")^{N_2}b"e_0 = 0$ and $b"x = 0$.

The next stage of the proof is to show that $L_-x = 0$. As follows from Eq. (8.14) and the definition of x,

$$L_-x = (a_1")^{N_1}(a_2")^{N_2}L_-e_0 - N_1(a_1")^{N_1-1}(a_2")^{N_2}b"e_0 \qquad (8.34)$$

We have already shown that $(a_2")^{N_2}b"e_0 = 0$, and therefore it is sufficient to prove that the first term in the r.h.s. of Eq. (8.34) is equal to zero. As follows from Eqs. (8.14) and (8.19), $a_2'L_-e_0 = b'e_0 = 0$, and L_-e_0 is the eigenvector of the operator h_2 with the eigenvalue $q_2 + 1$. Therefore $(a_2")^{N_2}L_-e_0 = 0$ and the proof is completed.

Let us assume for a moment that the eigenvalues of the operators h_1 and h_2 should be treated not as elements of F_p but as integers. Then, as follows from the consideration in the preceding section, if $f(q_j) \ll p$ (j=1,2) then one modular IR of the so(2,3) algebra corresponds to a standard positive energy IR in the region where the energy is positive and much less than p. At the same time, it corresponds to an IR with the negative energy in the region where the AdS energy is close to $4p$ but less than $4p$.

8.3 Massless particles, Dirac singletons and special IRs

Those cases can be considered by analogy with the massive one. The case of Dirac singletons is especially simple. As follows from Eqs. (8.24) and (8.26), if $(m_{AdS} = 1, s = 0)$ then the only possible value of k is $k = 0$ and the only possible values of n are $n = 0, 1$ while if $(m_{AdS} = 2, s = 1)$ then the only possible values of k are $k = 0, 1$ and the only possible value of n is $n = 0$. This result does not depend on the value of p and therefore it is valid in both, standard theory and GFQT. The only difference between standard and modular cases is that in the former $n_1, n_2 = 0, 1, ...\infty$ while in the latter the quantities n_1, n_2 are in the range defined by Eq. (8.28). In the literature, the IR with $(m_{AdS} = 2, s = 1)$ is called Di and the IR with $(m_{AdS} = 1, s = 0)$ is called Rac.

The singleton IRs are indeed exceptional since the value of n in them does not exceed 1 and therefore the impression is that singletons are two-dimensional objects, not three-dimensional ones as usual particles. However, the singleton IRs have been obtained in the so(2,3) theory without reducing the algebra. Dirac has titled his paper [109] "A Remarkable Representation of the 3 + 2 de Sitter Group". Below we argue that in GFQT the singleton IRs are even more remarkable than in standard theory.

First of all, as noted above, in standard theory there exist independent positive and negative IRs and the latter are associated with antiparticles. In particular, in standard theory there exist four singleton IRs - two IRs with positive energies and the corresponding IRs with negative energies, which can be called antisingletons. However, at the end of the preceding section we have proved that in GFQT one IR contains positive and negative energy states simultaneously. This proof can be applied to the singleton IRs without any changes. As a consequence, in the modular case there exist only two singleton IRs.

If $(m_{AdS} = 1, s = 0)$ then $q_1 = q_2 = 1/2$ and, as noted in the preceding section, in GFQT these relations should be treated as $q_1 = q_2 = (p+1)/2$. Analogously, if $(m_{AdS} = 2, s = 1)$ then $(q_1 = 3/2, q_2 = 1/2)$ and in GFQT $(q_1 = (p+3)/2, q_2 = (p+1)/2)$. Therefore the values of q_1 and q_2 for the singleton IRs are extremely large since they are of the order of $p/2$. As a consequence, the singleton IRs do not contain states where all the quantum numbers are much less than p. Since some of the quantum numbers are necessarily of the order of p, this is a natural explanation of the fact that singletons have not been observed. In addition, as follows from the discussion in Chap. 6 and Secs. 8.1 and 8.2, the fact that some quantum numbers are of the order of p implies that the singletons cannot be described in terms of the probabilistic interpretation.

Note also that if we consider the singleton IRs as modular analogs of negative energy IRs then the singleton IRs should be characterized either by $q_1 = q_2 = -1/2$ or by $q_1 = -3/2,\ q_2 = -1/2$. However, since in Galois fields $-1/2 = (p-1)/2$

and $-3/2 = (p-3)/2$, those values are very close to ones characterizing modular analogs of positive energy IRs. As a consequence, there is no approximation when singleton states can be characterized as particles or antiparticles.

The Rac IR contains only minimal $sp(2) \times sp(2)$ vectors with $h_1 = h_2 = (p+1)/2$ and $h_1 = h_2 = (p+3)/2$ while the Di IR contains only minimal $sp(2)\times sp(2)$ vectors with $h_1 = (p+3)/2,\ h_2 = (p+1)/2$ and $h_1 = (p+1)/2,\ h_2 = (p+3)/2$. Hence it easily follows from Eq. (8.7) that the dimensions of these IRs are equal to

$$Dim(Rac) = \frac{1}{2}(p^2+1) \quad Dim(Di) = \frac{1}{2}(p^2-1) \tag{8.35}$$

Consider now the massless case. Note first that when $q_2 = 1$, it follows from Eqs. (8.24) and (8.26) that $a(0) = 0$ and $b(0) = 0$. Therefore $A^{++}e_0 = A^{-+}e_0 = 0$ and if the definition $e(n,k) = (A^{++})^n(A^{-+})^k e_0$ is used for $(n = 0,1,...)$ and $(k = 0,1,...)$ then all the $e(n,k)$ will be the null elements.

We first consider the case when $s \neq 0$ and $s \neq p-1$. In that case we define $e(1,0)$ not as $A^{++}e_0$ but as $e(1,0) = [b"(h_1-1) - a_1"L_-]e_0$. A direct calculation using Eq. (8.14) shows that when $q_2 = 1$, this definition is legitimate since $e(1,0)$ is the minimal $sp(2)\times sp(2)$ vector with the eigenvalues of the operators h_1 and h_2 equal to $2+s$ and 2, respectively. With such a definition of $e(1,0)$, a direct calculation using Eqs. (8.9) and (8.14) gives $A^{--}e(1,0) = b'e(1,0) = s(s+1)e_0$ and therefore $e(1,0) \neq 0$. We now define $e(n,0)$ at $n \geq 1$ as $e(n,0) = (A^{++})^{n-1}e(1,0)$. Then Eq. (8.22) remains valid when $n \geq 1$. Since $A^{++}b'e(1,0) = s(s+1)A^{++}e_0 = 0$, Eq. (8.23) remains valid at $n = 1,2,...$ and $a(0) = 0$. Hence we get

$$a(n) = n(n+1)(n+s+1)(n+s) \quad (n \geq 1) \tag{8.36}$$

As a consequence, the maximal value of n in the modular case is $n_{max} = p-1-s$. This result has been obtained in Ref. [41].

For analogous reasons, we now cannot define $e(0,k)$ as $(A^{-+})^k e_0$. However, if we define $e(0,k) = (L_-)^k e_0$ then, as follows from the discussion at the end of Sec. 8.1, the elements $e(0,k)$ $(k = 0,1,...s)$ form a basis of the IR of the su(2) algebra with the spin s. Therefore the new definition of $e(0,k)$ is legitimate since $e(0,k)$ is the minimal $sp(2)\times sp(2)$ vector with the eigenvalues of the operators h_1 and h_2 equal to $1+s-k$ and $1+k$, respectively.

A direct calculation using Eqs. (8.9) and (8.14) gives that with the new definition of $e(0,k)$, $A^{--}A^{++}e(0,k) = b'A^{++}e(0,k) = 0$ and therefore $A^{++}e(0,k) = 0$. When $1 \leq k \leq s-1$, there is no way to obtain nonzero minimal $sp(2)\times sp(2)$ vectors with the eigenvalues of the operators h_1 and h_2 equal to $1+s-k+n$ and $1+k+n$, respectively, when $n > 0$. However, when $k = s$, such vectors can be obtained by analogy with the case $k = 0$. We define $e(1,s) = [b"(h_2-1) - a_2"L_+]e(0,s)$. Then a direct calculation gives $b'e(1,s) = s(s+1)e(0,s)$ and therefore $e(1,s) \neq 0$. We now define $e(n,s) = (A^{++})^{n-1}e(1,s)$ for $n \geq 1$. Then by analogy with the above discussion one can verify that if $A^{--}A^{++}e(n,s) = a(n)e(n,s)$ then $a(n)$ for $n \geq 1$ is again given by Eq. (8.36) and therefore in the modular case the maximal value of n is the same.

If $s = 0$ then the only possible value of k is $k = 0$ and for the vectors $e(n, 0)$ we have the same results as above. In particular, Eq. (8.36) is valid with $s = 0$. When $s = p - 1$, we can define $e(n, 0)$ and $e(n, s)$ as above but since $s + 1 = 0$ $(mod\ p)$, we get that $e(1, 0) = e(1, s) = 0$. This is in agreement with the above discussion since $n_{max} = 0$ when $s = p - 1$.

According to Standard Model (based on Poincare invariance), only massless Weyl particles can be fundamental elementary particles in Poincare invariant theory. Therefore a problem arises whether the above results can be treated as analogs of Weyl particles in standard and modular versions of AdS invariant theory. In view of the relation $P^\mu = M^{4\mu}/2R$ (see Sec. 1.3), the AdS mass m_{AdS} and the Poincare mass m are related as $m = m_{AdS}/2R$. Since $m_{AdS} = 2q_2 + s$, the corresponding Poincare mass will be zero when $R \to \infty$ not only when $q_2 = 1$ but when q_2 is any finite number. So a question arises why only the case $q_2 = 1$ is treated as massless. In Poincare invariant theory, Weyl particles are characterized not only by the condition that their mass is zero but also by the condition that they have a definite helicity. In standard case the minimum value of the AdS energy for massless IRs with positive energy is $E_{min} = 2 + s$ when $n = 0$. In contrast to the situation in Poincare invariant theory, where massless particles cannot be in the rest state, the massless particles in the AdS theory do have rest states and, as shown above, the value of the z projection of the spin in such states can be $-s, -s + 2, ...s$ as usual. However, we have shown that for any value of energy greater than E_{min}, when $n \neq 0$, the spin state is characterized only by helicity, which can take the values either s when $k = 0$ or $-s$ when $k = s$, i.e. we have the same result as in Poincare invariant theory. Note that in contrast to IRs of the Poincare and dS algebras, standard IRs describing particles in AdS invariant theory belong to the discrete series of IRs and the energy spectrum in them is discrete: $E = E_{min}, E_{min} + 2, ...\infty$. Therefore, strictly speaking, rest states do not have measure zero as in Poincare and dS invariant theories. Nevertheless, the probability that the energy is exactly E_{min} is extremely small and therefore the above results show that the case $q_2 = 1$ indeed describes AdS analogs of Weyl particles.

Consider now dimensions of massless IRs. If $s = 0$ then, as follows from the above results, there exist only minimal $sp(2) \times sp(2)$ vectors with $h_1 = h_2 = 1+n$, $n = 0, 1, ...p - 1$. Therefore, as follows from Eq. (8.7), the dimension of the massless IR with $s = 0$ equals

$$Dim(s = 0) = \sum_{n=0}^{p-1} (p - n)^2 = \frac{1}{6}p(p + 1)(2p + 1) \tag{8.37}$$

If $s = 1$, there exist only minimal $sp(2) \times sp(2)$ vectors with $(h_1 = 2 + n, h_2 = 1 + n)$

and $(h_1 = 1+n, h_2 = 2+n)$ where $n = 0, 1, ...p-2$. Therefore

$$Dim(s=1) = 2\sum_{n=0}^{p-2}(p-n)(p-n-1) = \frac{2}{3}p(p-1)(p+1) \tag{8.38}$$

If $s \geq 2$, there exist only minimal $sp(2)\times sp(2)$ vectors with $(h_1 = 1+s+n, h_2 = 1+n)$, $(h_1 = 1+n, h_2 = 1+s+n)$ where $n = 0, 1, ...p-s$ and the minimal $sp(2)\times sp(2)$ vectors with $(h_1 = 1+s-k, h_2 = 1+k)$ where $k = 1, ...s-1$. Therefore, as follows from Eq. (8.7)

$$Dim(s \geq 2) = 2\sum_{n=0}^{p-s}(p-n)(p-n-s) + \sum_{k=1}^{s-1}(p-k)(p-s+k) =$$
$$\frac{p}{3}(2p^2 - 3s^2 + 1) + \frac{1}{2}s(s-1)(s+1) \tag{8.39}$$

As noted in Sec. 8.2, the cases of special IRs are such either q_1 and q_2 are represented by the numbers $0, 1, ..p-1$ and $q_1 < q_2$ or in the case of special singletons, $q_1, q_2 = 2, ...p-1$ but $(q_1 + q_2)$ $(mod\ p)$ is one of the numbers (0,1,2). For example, $(q_1 = (p+1)/2, q_2 = (p-1)/2)$ is a special singleton with $(m_{AdS} = 0, s = 1)$, $(q_1 = (p+3)/2, q_2 = (p-1)/2)$ is a special singleton with $(m_{AdS} = 1, s = 2)$ etc. These cases can be investigated by analogy with massive IRs in Sec. 8.2. For reasons given in Sec. 8.10 and Chap. 9, among singleton IRs we will consider in detail only the Dirac singletons. Then we will see that the only special IRs taking part in the decomposition of the tensor product of the Dirac singletons are those with $q_1 = 0$. Then $s = p - q_2$. If $q_2 = 2, 3, ...p-1$ then, as follows from Eq. (8.24), the quantum number n can take only the value $n = 0$. If $q_2 = 1$ then the special IR can also be treated as the massless IR with $s = p-1$. As noted above, in this case the quantity n also can take only the value $n = 0$. Let $Dim(q_1, q_2)$ be the dimension of the IR characterized by q_1 and q_2. Then, as follows from Eq. (8.7)

$$Dim(0, q_2) = \sum_{k=0}^{p-q_2}(1+p-q_2-k)(1+k) = (1+p-q_2)^2 + \frac{1}{2}(p-q_2)^2(1+p-q_2) \tag{8.40}$$

8.4 Matrix elements of representation operators

In what follows, we will discuss the massive case but the same results are valid in the singleton and massless cases. The matrix elements of the operator A are defined as

$$Ae(n_1n_2nk) = \sum_{n_1'n_2'n'k'} A(n_1'n_2'n'k'; n_1n_2nk)e(n_1'n_2'n'k') \tag{8.41}$$

where the sum is taken over all possible values of $(n'_1n'_2n'k')$. One can explicitly calculate matrix elements for all the representation operators and the results are:

$$
\begin{aligned}
h_1e(n_1n_2nk) &= [Q_1(n,k)+2n_1]e(n_1n_2nk)\\
h_2e(n_1n_2nk) &= [Q_2(n,k)+2n_2]e(n_1n_2nk)
\end{aligned} \tag{8.42}
$$

$$
\begin{aligned}
a'_1e(n_1n_2nk) &= n_1[Q_1(n,k)+n_1-1]e(n_1-1,n_2nk)\\
a_1"e(n_1n_2nk) &= e(n_1+1,n_2nk)\\
a'_2e(n_1n_2nk) &= n_2[Q_2(n,k)+n_2-1]e(n_1,n_2-1,nk)\\
a_2"e(n_1n_2nk) &= e(n_1,n_2+1,nk)
\end{aligned} \tag{8.43}
$$

$$
\begin{gathered}
b"e(n_1n_2nk) = \{[Q_1(n,k)-1][Q_2(n,k)-1]\}^{-1}\\
[k(s+1-k)(q_1-k-1)(q_2+k-2)e(n_1,n_2+1,n,k-1)+\\
n(m_{AdS}+n-3)(q_1+n-1)(q_2+n-2)e(n_1+1,n_2+1,n-1,k)+\\
e(n_1,n_2,n+1,k)+e(n_1+1,n_2,n,k+1)]
\end{gathered} \tag{8.44}
$$

$$
\begin{gathered}
b'e(n_1n_2nk) = \{[Q_1(n,k)-1][Q_2(n,k)-1]\}^{-1}[n(m_{AdS}+n-3)\\
(q_1+n-1)(q_2+n-2)(q_1+n-k+n_1-1)(q_2+n+k+n_2-1)\\
e(n_1n_2,n-1,k)+n_2(q_1+n-k+n_1-1)e(n_1,n_2-1,n,k+1)+\\
n_1(q_2+n+k+n_2-1)k(s+1-k)(q_1-k-1)(q_2+k-2)\\
e(n_1-1,n_2,n,k-1)+n_1n_2e(n_1-1,n_2-1,n+1,k)]
\end{gathered} \tag{8.45}
$$

$$
\begin{gathered}
L_+e(n_1n_2nk) = \{[Q_1(n,k)-1][Q_2(n,k)-1]\}^{-1}\{(q_2+n+k+n_2-1)\\
[k(s+1-k)(q_1-k-1)(q_2+k-2)e(n_1n_2n,k-1)+\\
n(m_{AdS}+n-3)(q_1+n-1)(q_2+n-2)e(n_1+1,n_2,n-1,k)]+\\
n_2[e(n_1,n_2-1,n+1,k)+e(n_1+1,n_2-1,n,k+1)]\}
\end{gathered} \tag{8.46}
$$

$$
\begin{gathered}
L_-e(n_1n_2nk) = \{[Q_1(n,k)-1][Q_2(n,k)-1]\}^{-1}\{n_1[k(s+1-k)\\
(q_1-k-1)(q_2+k-2)e(n_1-1,n_2n,k-1)+e(n_1-1,n_2,n+1,k)]\\
+(q_1+n-k+n_1-1)[e(n_1n_2n,k+1)+n(m_{AdS}+n-3)\\
(q_1+n-1)(q_2+n-2)e(n_1,n_2+1,n-1,k)]\}
\end{gathered} \tag{8.47}
$$

We will always use a convention that $e(n_1n_2nk)$ is a null vector if some of the numbers (n_1n_2nk) are not in the range described above.

The important difference between standard and modular IRs is that in the latter the trace of each representation operator is equal to zero while in the former this is obviously not the case (for example, the energy operator is positive definite for

IRs defined by Eq. (8.19) and negative definite for IRs defined by Eq. (8.32)). For the operators $(a'_j, a_j", L_\pm, b', b")$ the validity of this statement is clear immediately: since they necessarily change one of the quantum numbers $(n_1 n_2 n k)$, they do not contain nonzero diagonal elements at all. The proof for the diagonal operators h_1 and h_2 follows. For each IR of the sp(2) algebra with the "minimal weight" q_0 and the dimension $N+1$, the eigenvalues of the operator h are $(q_0, q_0+2, ...q_0+2N)$. The sum of these eigenvalues equals zero in F_p since $q_0+N=0$ in F_p (see Sec. 8.1). Therefore we conclude that for any representation operator A

$$\sum_{n_1 n_2 n k} A(n_1 n_2 n k, n_1 n_2 n k) = 0 \qquad (8.48)$$

This property is very important for investigating a new symmetry between particles and antiparticles in the GFQT which is discussed in the subsequent section.

8.5 Quantization and AB symmetry

Let us first consider how the Fock space can be defined in standard theory. As shown in Sec. 8.2, in the AdS case (in contrast to the situation in the dS one) IRs with positive and negative energies are fully independent. Let (n_1, n_2, n, k) be the set of all quantum numbers characterizing basis vectors of the IR and $a(n_1 n_2 n k)$ be the operator of particle annihilation in the state described by the vector $e(n_1 n_2 n k)$. Then the adjoint operator $a(n_1 n_2 n k)^*$ has the meaning of particle creation in that state. Since we do not normalize the states $e(n_1 n_2 n k)$ to one, we require that the operators $a(n_1 n_2 n k)$ and $a(n_1 n_2 n k)^*$ should satisfy either the anticommutation relations

$$\{a(n_1 n_2 n k), a(n'_1 n'_2 n' k')^*\} = Norm(n_1 n_2 n k)\delta_{n_1 n'_1}\delta_{n_2 n'_2}\delta_{nn'}\delta_{kk'} \qquad (8.49)$$

or the commutation relations

$$[a(n_1 n_2 n k), a(n'_1 n'_2 n' k')^*] = Norm(n_1 n_2 n k)\delta_{n_1 n'_1}\delta_{n_2 n'_2}\delta_{nn'}\delta_{kk'} \qquad (8.50)$$

A problem arises that in the case of negative energy IRs the operators $a(n_1 n_2 n k)$ and $a(n_1 n_2 n k)^*$ have the meaning of the annihilation and creation operators, respectively, for the states with negative energies and hence a question arises of whether such operators are physical. An analogous problem for the dS case has been discussed in Sec. 3.5. One might think that since in the AdS case IRs with positive and negative energies are fully independent, we can simply declare IRs with negative energies unphysical and consider only IRs with positive energies. However, in QFT one cannot get rid of negative energy IRs since here positive and negative energy IRs are combined together into a field satisfying a local covariant equation. For example, the Dirac field combines together positive and negative energy IRs into the Dirac field satisfying the Dirac equation.

For combining two IRs with positive and negative energies together, one can introduce a new quantum number ϵ which will distinguish IRs with positive and negative energies; for example $\epsilon = \pm 1$ for the positive and negative energy IRs, respectively. Then we have a set of operators $a(n_1n_2nk,\epsilon)$ and $a(n_1n_2nk,\epsilon)^*$ such that by analogy with Eq. (8.49)

$$\{a(n_1n_2nk,\epsilon), a(n_1'n_2'n'k',\epsilon')^*\} = Norm(n_1n_2nk)\delta_{n_1n_1'}\delta_{n_2n_2'}\delta_{nn'}\delta_{kk'}\delta_{\epsilon\epsilon'} \tag{8.51}$$

and analogously in the case of commutators. The vacuum state $\tilde{\Phi}_{vac}$ can be defined by the condition

$$a(n_1n_2nk,\epsilon)\tilde{\Phi}_{vac} = 0 \quad \forall(n_1,n_2,n,k,\epsilon) \tag{8.52}$$

As follows from Eqs. (8.15) and (8.42), the secondly quantized energy operator has the form

$$M^{04} = \sum_{n_1n_2nk,\epsilon} \epsilon[m_{AdS} + 2(n+n_1+n_2)]a(n_1n_2nk,\epsilon)^* a(n_1n_2nk,\epsilon) \tag{8.53}$$

and hence we have to solve the problem of the physical interpretation of the operators $a(n_1n_2nk,-1)$ and $a(n_1n_2nk,-1)^*$. The two well-known ways of solving this problem follow.

In the spirit of Dirac's hole theory, one can define the new physical vacuum

$$\Phi_{vac} = \prod_{n_1n_2nk} a(n_1n_2nk,-1)^*\tilde{\Phi}_{vac} \tag{8.54}$$

Then in the case of anticommutators each operator $a(n_1n_2nk,-1)$ creates a hole with a negative energy and the corresponding operator $a(n_1n_2nk,-1)^*$ annihilates this hole. Hence the operators $a(n_1n_2nk,-1)^*$ can now be treated as the annihilation operators of states with positive energies and the operators $a(n_1n_2nk,-1)$ — as the creation operators of states with positive energies. A problem with such a treatment is that Φ_{vac} is the eigenstate of the operator M^{04} with the eigenvalue

$$\mathcal{E}_{vac} = -\sum_{n_1n_2nk}[m_{AdS} + 2(n+n_1+n_2)] \tag{8.55}$$

This is an infinite negative value and in quantum gravity a vacuum with an infinite energy is treated as unacceptable.

Another approach is that we consider only quantum numbers describing IRs with positive energies and, in addition to the operators $a(n_1n_2nk) = a(n_1n_2nk,1)$ and $a(n_1n_2nk)^* = a(n_1n_2nk,1)^*$, introduce new operators $b(n_1n_2nk)$ and $b(n_1n_2nk)^*$ instead of the operators $a(n_1n_2nk,-1)$ and $a(n_1n_2nk,-1)^*$ such that $b(n_1n_2nk)$ is proportional to $a(n_1n_2nk,-1)^*$ and $b(n_1n_2nk)^*$ is proportional to $a(n_1n_2nk,-1)$. Then the b-operators are treated as the annihilation operators of antiparticles with positive energies and the b^* operators — as the creation operators of antiparticles with positive energies. By analogy with Eqs. (8.49) and (8.50), they should satisfy the relations

$$\{b(n_1n_2nk), b(n'_1n'_2n'k')^*\} = Norm(n_1n_2nk)\delta_{n_1n'_1}\delta_{n_2n'_2}\delta_{nn'}\delta_{kk'} \tag{8.56}$$

$$[b(n_1n_2nk), b(n'_1n'_2n'k')^*] = Norm(n_1n_2nk)\delta_{n_1n'_1}\delta_{n_2n'_2}\delta_{nn'}\delta_{kk'} \tag{8.57}$$

for anticommutation or commutation relations, respectively. In this case it is assumed that in the case of anticommutation relations all the operators (a, a^*) anticommute with all the operators (b, b^*) while in the case of commutation relations they commute with each other. It is also assumed that the vacuum vector Φ_0 should satisfy the conditions

$$a(n_1n_2nk)\Phi_0 = b(n_1n_2nk)\Phi_0 = 0 \quad \forall\ n_1, n_2, n, k \tag{8.58}$$

In QFT the second possibility is treated as more physical than that analogous to Dirac's hole theory.

The Fock space in standard theory can now be defined as a linear combination of all elements obtained by the action of the operators (a^*, b^*) on the vacuum vector, and the problem of second quantization of representation operators can be formulated as follows. Let $(A_1, A_2....A_n)$ be representation operators describing IR of the AdS algebra. One should replace them by operators acting in the Fock space such that the commutation relations between their images in the Fock space are the same as for original operators (in other words, we should have a homomorphism of Lie algebras of operators acting in the space of IR and in the Fock space). We can also require that our map should be compatible with the Hermitian conjugation in both spaces. It is easy to verify that a possible solution satisfying all the requirements is as follows. Taking into account the fact that the matrix elements satisfy the proper commutation relations, the operators A_i in the quantized form

$$\begin{aligned} A_i = \sum A_i(n'_1n'_2n'k', n_1n_2nk)[a(n'_1n'_2n'k')^*a(n_1n_2nk) + \\ b(n'_1n'_2n'k')^*b(n_1n_2nk)]/Norm(n_1n_2nk) \end{aligned} \tag{8.59}$$

satisfy the commutation relations (8.9,8.13,8.14). Here the sum is taken over all the possible quantum numbers $(n'_1, n'_2, n', k', n_1, n_2, n, k)$. We will not use special notations for operators in the Fock space since in each case it will be clear whether the operator in question acts in the space of IR or in the Fock space.

A well-known problem in standard theory is that the quantization procedure does not define the order of the annihilation and creation operators uniquely. For example, another possible solution is

$$\begin{aligned} A_i = \mp \sum A_i(n'_1n'_2n'k', n_1n_2nk)[a(n_1n_2nk)a(n'_1n'_2n'k')^* + \\ b(n_1n_2nk)b(n'_1n'_2n'k')^*]/Norm(n_1n_2nk) \end{aligned} \tag{8.60}$$

for anticommutation and commutation relations, respectively. The solutions (8.59) and (8.60) are different since the energy operators M^{04} in these expressions differ by an infinite constant. In standard theory the solution (8.59) is selected by imposing an additional requirement that all operators should be written in the normal form where annihilation operators precede creation ones. Then the vacuum has zero energy and Eq. (8.60) should be rejected. Such a requirement does not follow from the theory. Ideally there should be a procedure which correctly defines the order of operators from first principles.

In standard theory there also exist neutral particles. In that case there is no need to have two independent sets of operators (a,a^*) and (b,b^*), and Eq. (8.59) should be written without the (b,b^*) operators. The problem of neutral particles in GFQT is discussed in Sec. 8.9.

We now proceed to quantization in the modular case. The results of Sec. 8.2 show that one modular IR corresponds to two standard IRs with the positive and negative energies, respectively. This indicates to a possibility that one modular IR describes a particle and its antiparticle simultaneously. However, we don't know yet what should be treated as a particle and its antiparticle in the modular case. We have a description of an object such that (n_1n_2nk) is the full set of its quantum numbers which take the values described in the preceding section.

We now assume that $a(n_1n_2nk)$ in GFQT is the operator describing annihilation of the object with the quantum numbers (n_1n_2nk) regardless of whether the numbers are physical or nonphysical. Analogously $a(n_1n_2nk)^*$ describes creation of the object with the quantum numbers (n_1n_2nk). If these operators anticommute then they satisfy Eq. (8.49) while if they commute then they satisfy Eq. (8.50). Then, by analogy with standard case, the operators

$$A_i = \sum A_i(n_1'n_2'n'k', n_1n_2nk)a(n_1'n_2'n'k')^*a(n_1n_2nk)/Norm(n_1n_2nk) \qquad (8.61)$$

satisfy the commutation relations (8.9,8.13,8.14). In this expression the sum is taken over all possible values of the quantum numbers in the modular case.

In the modular case the solution can be taken not only as in Eq. (8.61) but also as

$$A_i = \mp\sum A_i(n_1'n_2'n'k', n_1n_2nk)a(n_1n_2nk)a(n_1'n_2'n'k')^*/Norm(n_1n_2nk) \qquad (8.62)$$

for the cases of anticommutators and commutators, respectively. However, as follows from Eqs. (8.48-8.50), the solutions (8.61) and (8.62) are the same. Therefore in the modular case there is no need to impose an artificial requirement that all operators should be written in the normal form.

The problem with the treatment of the (a,a^*) operators follows. When the values of (n_1n_2n) are much less than p, the modular IR corresponds to standard positive energy IR and therefore the (a,a^*) operator can be treated as those describing the particle annihilation and creation, respectively. However, when the AdS energy is negative, the operators $a(n_1n_2nk)$ and $a(n_1n_2nk)^*$ become unphysical since they describe annihilation and creation, respectively, in the unphysical region of negative energies.

Let us recall that at any fixed values of n and k, the quantities n_1 and n_2 can take only the values described in Eq. (8.28) and the eigenvalues of the operators h_1 and h_2 are given by $Q_1(n,k)+2n_1$ and $Q_2(n,k)+2n_2$, respectively. As follows from Eq. (8.7) and the results of Sec. 8.2, the first IR of the sp(2) algebra has the dimension $N_1(n,k)+1$ and the second IR has the dimension $N_2(n,k)+1$. If $n_1 = N_1(n,k)$ then it follows from Eq. (8.28) that the first eigenvalue is equal to $-Q_1(n,k)$ in F_p, and if $n_2 = N_2(n,k)$ then the second eigenvalue is equal to $-Q_2(n,k)$ in F_p. We use $\tilde{n}_1$ to denote $N_1(n,k)-n_1$ and $\tilde{n}_2$ to denote $N_2(n,k)-n_2$. Then it follows from Eq. (8.28) that $e(\tilde{n}_1\tilde{n}_2nk)$ is the eigenvector of the operator h_1 with the eigenvalue $-(Q_1(n,k)+2n_1)$ and the eigenvector of the operator h_2 with the eigenvalue $-(Q_2(n,k)+2n_2)$.

As noted above, standard theory involves the idea that creation of the antiparticle with positive energy can be treated as annihilation of the corresponding particle with negative energy and annihilation of the antiparticle with positive energy can be treated as creation of the corresponding particle with negative energy. In GFQT we also can define the operators $b(n_1n_2nk)$ and $b(n_1n_2nk)^*$ in such a way that they will replace the (a,a^*) operators if the quantum numbers are unphysical. In addition, if the values of (n_1n_2n) are much less than p, the operators $b(n_1n_2nk)$ and $b(n_1n_2nk)^*$ should be interpreted as physical operators describing annihilation and creation of antiparticles, respectively.

In GFQT the (b,b^*) operators cannot be independent of the (a,a^*) operators since the latter are defined for all possible quantum numbers. Therefore the (b,b^*) operators should be expressed in terms of the (a,a^*) ones. We can implement the above idea if the operator $b(n_1n_2nk)$ is defined in such a way that it is proportional to $a(\tilde{n}_1,\tilde{n}_2,n,k)^*$ and hence $b(n_1n_2nk)^*$ is proportional to $a(\tilde{n}_1,\tilde{n}_2,n,k)$.

Since Eq. (8.31) should now be considered in F_p, it follows from the well-known Wilson theorem $(p-1)! = -1$ in F_p (see e.g. [98]) that

$$F(n_1n_2nk)F(\tilde{n}_1\tilde{n}_2nk) = (-1)^s \tag{8.63}$$

We now define the b-operators as

$$a(n_1n_2nk)^* = \eta(n_1n_2nk)b(\tilde{n}_1\tilde{n}_2nk)/F(\tilde{n}_1\tilde{n}_2nk) \tag{8.64}$$

where $\eta(n_1n_2nk)$ is some function. As a consequence,

$$\begin{aligned}
a(n_1n_2nk) &= \bar{\eta}(n_1n_2nk)b(\tilde{n}_1\tilde{n}_2nk)^*/F(\tilde{n}_1\tilde{n}_2nk)\\
b(n_1n_2nk)^* &= a(\tilde{n}_1\tilde{n}_2nk)F(n_1n_2nk)/\bar{\eta}(\tilde{n}_1\tilde{n}_2nk)\\
b(n_1n_2nk) &= a(\tilde{n}_1\tilde{n}_2nk)^*F(n_1n_2nk)/\eta(\tilde{n}_1\tilde{n}_2nk)
\end{aligned} \tag{8.65}$$

Equations (8.64) and (8.65) define a relation between the sets (a,a^*) and (b,b^*). Although our motivation was to replace the (a,a^*) operators by the (b,b^*)

ones only for the nonphysical values of the quantum numbers, we can consider this definition for all the values of (n_1n_2nk). The transformation described by Eqs. (8.64) and (8.65) can also be treated as a special case of the Bogolubov transformation discussed in a wide literature on many-body theory (see e.g., Chap. 10 in Reference [88] and references therein).

We have not discussed yet what exact definition of the physical and nonphysical quantum numbers should be. This problem will be discussed in Sec. 8.6. However, one might accept

Physical-nonphysical states assumption: Each set of quantum numbers (n_1n_2nk) is either physical or unphysical. If it is physical then the set $(\tilde{n}_1\tilde{n}_2nk)$ is unphysical and vice versa.

With this assumption we can conclude from Eqs. (8.64) and (8.65) that if some operator a is physical then the corresponding operator b^* is unphysical and vice versa while if some operator a^* is physical then the corresponding operator b is unphysical and vice versa.

We have no ground to think that the set of the (a, a^*) operators is more fundamental than the set of the (b, b^*) operators and vice versa. Therefore the question arises whether the (b, b^*) operators satisfy the relations (8.50) or (8.56) in the case of anticommutation or commutation relations, respectively and whether the operators A_i (see Eq. (8.61)) have the same form in terms of the (a, a^*) and (b, b^*) operators. In other words, if the (a, a^*) operators in Eq. (8.61) are expressed in terms of the (b, b^*) ones then the problem arises whether

$$A_i = \sum A_i(n_1'n_2'n'k', n_1n_2nk)b(n_1'n_2'n'k')^*b(n_1n_2nk)/Norm(n_1n_2nk) \qquad (8.66)$$

is valid. It is natural to accept the following

Definition of the AB symmetry: If the (b, b^*) operators satisfy Eq. (8.56) in the case of anticommutators or Eq. (8.57) in the case of commutators and all the representation operators (8.61) in terms of the (b, b^*) operators have the form (8.66) then it is said that the AB symmetry is satisfied.

To prove the AB symmetry we will first investigate whether Eqs. (8.56) and (8.57) follow from Eqs. (8.49) and (8.50), respectively. As follows from Eqs. (8.63-8.65), Eq. (8.56) follows from Eq. (8.49) if

$$\eta(n_1n_2nk)\bar{\eta}(n_1, n_2, nk) = (-1)^s \qquad (8.67)$$

while Eq. (8.57) follows from Eq. (8.50) if

$$\eta(n_1n_2nk)\bar{\eta}(n_1, n_2, nk) = (-1)^{s+1} \qquad (8.68)$$

We now represent $\eta(n_1n_2nk)$ in the form

$$\eta(n_1n_2nk) = \alpha f(n_1n_2nk) \qquad (8.69)$$

where $f(n_1n_2nk)$ should satisfy the condition

$$f(n_1n_2nk)\bar{f}(n_1,n_2,nk)=1 \qquad (8.70)$$

Then α should be such that

$$\alpha\bar{\alpha}=\pm(-1)^s \qquad (8.71)$$

where the plus sign refers to anticommutators and the minus sign to commutators, respectively. If the normal spin-statistics connection is valid, i.e. we have anticommutators for odd values of s and commutators for even ones then the r.h.s. of Eq. (8.71) equals -1 while in the opposite case it equals 1. In Sec. 8.9, Eq. (8.71) is discussed in detail and for now we assume that solutions of this relation exist.

A direct calculation using the explicit expressions (8.42-8.47) for the matrix elements shows that if $\eta(n_1n_2nk)$ is given by Eq. (8.69) and

$$f(n_1n_2nk)=(-1)^{n_1+n_2+n} \qquad (8.72)$$

then the AB symmetry is valid regardless of whether the normal spin-statistics connection is valid or not.

8.6 Physical and nonphysical states

The operator $a(n_1n_2nk)$ can be the physical annihilation operator only if it annihilates the vacuum vector Φ_0. Then if the operators $a(n_1n_2nk)$ and $a(n_1n_2nk)^*$ satisfy the relations (8.49) or (8.50), the vector $a(n_1n_2nk)^*\Phi_0$ has the meaning of the one-particle state. The same can be said about the operators $b(n_1n_2nk)$ and $b(n_1n_2nk)^*$. For these reasons in standard theory it is required that the vacuum vector should satisfy the conditions (8.58). Then the elements

$$\Phi_+(n_1n_2nk)=a(n_1n_2nk)^*\Phi_0,\quad \Phi_-(n_1n_2nk)=b(n_1n_2nk)^*\Phi_0 \qquad (8.73)$$

have the meaning of one-particle states for particles and antiparticles, respectively.

However, if one requires the condition (8.58) in GFQT, then it is obvious from Eqs. (8.64) and (8.65) that the elements defined by Eq. (8.73) are null vectors. Note that in standard approach the AdS energy is always greater than m_{AdS} while in GFQT the AdS energy is not positive definite. We can therefore try to modify Eq. (8.58) as follows. Suppose that *Physical-nonphysical states assumption* (see Sec. 8.5) can be substantiated. Then we can break the set of elements (n_1n_2nk) into two nonintersecting parts with the same number of elements, S_+ and S_-, such that if $(n_1n_2nk)\in S_+$ then $(\tilde{n}_1\tilde{n}_2nk)\in S_-$ and vice versa. Then, instead of the condition (8.58) we require

$$a(n_1n_2nk)\Phi_0=b(n_1n_2nk)\Phi_0=0\quad \forall\ (n_1,n_2,n,k)\in S_+ \qquad (8.74)$$

In that case the elements defined by Eq. (8.73) will indeed have the meaning of one-particle states for $(n_1n_2nk) \in S_+$.

It is clear that if we wish to work with the full set of elements (n_1n_2nk) then, as follows from Eqs. (8.64) and (8.65), the operators (b, b^*) are redundant and we can work only with the operators (a, a^*). However, if one works with the both sets, (a, a^*) and (b, b^*) then such operators can be independent of each other only for a half of the elements (n_1n_2nk).

Regardless of how the sets S_+ and S_- are defined, the *Physical-nonphysical states assumption* cannot be consistent if there exist quantum numbers (n_1n_2nk) such that $n_1 = \tilde{n}_1$ and $n_2 = \tilde{n}_2$. Indeed, in that case the sets (n_1n_2nk) and $(\tilde{n}_1\tilde{n}_2nk)$ are the same what contradicts the assumption that each set (n_1n_2nk) belongs either to S_+ or S_-.

Since the replacements $n_1 \to \tilde{n}_1$ and $n_2 \to \tilde{n}_2$ change the signs of the eigenvalues of the h_1 and h_2 operators (see Sec. 8.5), the condition that that $n_1 = \tilde{n}_1$ and $n_2 = \tilde{n}_2$ should be valid simultaneously implies that the eigenvalues of the operators h_1 and h_2 should be equal to zero simultaneously. Recall that (see Sec. 8.1) if one considers IR of the sp(2) algebra and treats the eigenvalues of the diagonal operator h not as elements of F_p but as integers, then they take the values of $q_0, q_0 + 2, ...2p - q_0 - 2, 2p - q_0$. Therefore the eigenvalue is equal to zero in F_p only if it is equal to p when considered as an integer. Since $m_{AdS} = q_1 + q_2$ and the AdS energy is $E = h_1 + h_2$, the above situation can take place only if the energy considered as an integer is equal to 2p. It now follows from Eq. (8.15) that the energy can be equal to $2p$ only if m_{AdS} is even. Since $s = q_1 - q_2$, we conclude that m_{AdS} can be even if and only if s is even. In that case we will necessarily have quantum numbers (n_1n_2nk) such that the sets (n_1n_2nk) and $(\tilde{n}_1\tilde{n}_2nk)$ are the same and therefore the *Physical-nonphysical states assumption* is not valid. On the other hand, if s is odd (*i.e.* half-integer in the usual units) then there are no quantum numbers (n_1n_2nk) such that the sets (n_1n_2nk) and $(\tilde{n}_1\tilde{n}_2nk)$ are the same.

Our conclusion is as follows: *If the separation of states should be valid for any quantum numbers then the spin s should be necessarily odd.* In other words, if the notion of particles and antiparticles is absolute then elementary particles can have only a half-integer spin in the usual units.

In view of the above observations it seems natural to implement the *Physical-nonphysical states assumption* as follows. *If the quantum numbers (n_1n_2nk) are such that $m_{AdS} + 2(n_1 + n_2 + n) < 2p$ then the corresponding state is physical and belongs to S_+, otherwise the state is unphysical and belongs to S_-.* However, one cannot guarantee that there are no other reasonable implementations.

8.7 AdS symmetry breaking

In view of the above discussion, our next goal is the following. We should take the operators in the form (8.61) and replace the (a, a^*) operators by the (b, b^*) ones only if $(n_1n_2nk) \in S_-$. Then a question arises whether we will obtain the standard result (8.59) where a sum is taken only over values of $(n_1n_2nk) \in S_+$. The fact that we have proved the AB symmetry does not guarantee that this is the case since the AB symmetry implies that the replacement has been made for all the quantum numbers, not only half of them. However, the derivation of the AB symmetry shows that for the contribution of such quantum numbers that $(n_1n_2nk) \in S_+$ and $(n_1'n_2'n'k') \in S_+$ we will indeed have the result (8.59) up to some constants. This derivation also guarantees that if we consider the action of the operators on states described by physical quantum numbers and the result of the action also is a state described by physical quantum numbers then on such states the correct commutation relations are satisfied. A problem arises whether they will be satisfied for transitions between physical and nonphysical quantum numbers.

Let $A(a_1^{'})$ be the secondly quantized operator corresponding to $a_1^{'}$ and $A(a_1^{"})$ be the secondly quantized operator corresponding to $a_1^{"}$. Consider the action of these operators on the state $\Phi = a(n_1n_2nk)^*\Phi_0$ such that $(n_1n_2nk) \in S_+$ but $(n_1+1, n_2nk) \in S_-$. As follows from Eqs. (8.13) and (8.42), we should have

$$[A(a_1^{'}), A(a_1^{"})]\Phi = [Q_1(n,k) + 2n_1]\Phi \tag{8.75}$$

As follows from Eqs. (8.43) and (8.64), $A(a_1^{"})\Phi = a(n_1+1, n_2nk)^*\Phi_0$. Since $(n_1 + 1, n_2nk) \in S_-$, we should replace $a(n_1+1, n_2nk)^*$ by an operator proportional to $b(\tilde{n}_1 - 1, \tilde{n}_2nk)$ and then, as follows from Eq. (8.58), $A(a_1^{"})\Phi = 0$. Now, by using Eqs. (8.43) and (8.64), we get

$$[A(a_1^{'}), A(a_1^{"})]\Phi = n_1[Q_1(n,k) + n_1 - 1]\Phi \tag{8.76}$$

Equations (8.75) and (8.76) are incompatible with each other and we conclude that our procedure breaks the AdS symmetry for transitions between physical and nonphysical states.

We conclude that if, by analogy with standard theory, one wishes to interpret modular IRs of the dS algebra in terms of particles and antiparticles then the commutation relations of the dS algebra will be broken. This does not mean that such a possibility contradicts the existing knowledge since they will be broken only at extremely high dS energies of the order of p. At the same time, a possible point of view is that since we started from the symmetry algebra and treat the conditions (4.1) as a must, we should not sacrifice symmetry because we don't know other ways of interpreting IRs. So we have the following dilemma: *Either the notions of particles and antiparticles are always valid and the commutation relations (4.1) are broken at very large AdS energies of the order of p or the commutation relations (4.1) are not broken and the notion of a particle and its antiparticle is only approximate.* In the latter case such additive quantum numbers as the electric charge and the baryon and lepton quantum numbers can be only approximately conserved.

8.8 Dirac vacuum energy problem

The Dirac vacuum energy problem is discussed in practically every textbook on QFT. In its simplified form it can be described as follows. Suppose that the energy spectrum is discrete and n is the quantum number enumerating the states. Let $E(n)$ be the energy in the state n. Consider the electron-positron field. As a result of quantization one gets for the energy operator

$$E = \sum_n E(n)[a(n)^* a(n) - b(n)b(n)^*] \tag{8.77}$$

where $a(n)$ is the operator of electron annihilation in the state n, $a(n)^*$ is the operator of electron creation in the state n, $b(n)$ is the operator of positron annihilation in the state n and $b(n)^*$ is the operator of positron creation in the state n. It follows from this expression that only anticommutation relations are possible since otherwise the energy of positrons will be negative. However, if anticommutation relations are assumed, it follows from Eq. (8.77) that

$$E = \{\sum_n E(n)[a(n)^* a(n) + b(n)^* b(n)]\} + E_0 \tag{8.78}$$

where E_0 is some infinite negative constant. Its presence was a motivation for developing Dirac's hole theory. In the modern approach it is usually required that the vacuum energy should be zero. This can be obtained by assuming that all operators should be written in the normal form. However, this requirement is not quite consistent since the result of quantization is Eq. (8.77) where the positron operators are not written in that form (see also the discussion in Sec. 8.5).

Consider now the AdS energy operator $M^{04} = h_1 + h_2$ in GFQT. As follows from Eqs. (8.42) and (8.62)

$$M^{04} = \sum [m_{AdS} + 2(n_1 + n_2 + n)] a(n_1 n_2 n k)^* a(n_1 n_2 n k)/Norm(n_1 n_2 n k) \tag{8.79}$$

where the sum is taken over all possible quantum numbers $(n_1 n_2 n k)$. As noted in the preceding section, the two most well-known ways of solving the problem of negative energies are either in the spirit of Dirac's hole theory or by using the notion of antiparticles.

Consider first the second possibility. Then as follows from Eqs. (8.63-8.65) and (8.69-8.71)

$$\begin{aligned} M^{04} = \{\textstyle\sum_{S_+} [m + 2(n_1 + n_2 + n)][a(n_1 n_2 n k)^* a(n_1 n_2 n k) + \\ b(n_1 n_2 n k)^* b(n_1 n_2 n k)]/Norm(n_1 n_2 n k)\} + \mathcal{E}_{vac} \end{aligned} \tag{8.80}$$

where the vacuum energy is given by

$$\mathcal{E}_{vac} = \mp \sum_{S_+} [m_{AdS} + 2(n_1 + n_2 + n)] \tag{8.81}$$

in the cases when the (b, b^*) operators anticommute and commute, respectively. For definiteness, we consider the case when the operators anticommute and therefore the sum in the r.h.s. of Eq. (8.81) is taken with the minus sign.

In the approach similar to Dirac's hole theory one can define a new vacuum in GFQT by analogy with Eq. (8.54):

$$\Phi_{vac} = \prod_{S_-} a(n_1 n_2 n k, -1)^* \Phi_0 \tag{8.82}$$

where the product is taken over all the quantum numbers belonging to S_-. Then, as follows from the definition of the sets S_+ and S_-, this vacuum will be the eigenstate of the operator M^{04} with the the same eigenvalue $\mathcal{E}_{vac}$ as that given by Eq. (8.81) with the minus side in the r.h.s.

As noted in the dilemma at the end of the preceding section, in the approach involving the b operators the commutation relations (4.1) are necessarily broken at very large values of the AdS energy while in the approach similar to Dirac's hole theory there is no need to introduce the b operators. In modern QFT the approach with the b operators is treated as preferable since the condition $\mathcal{E}_{vac} = 0$ can be satisfied by imposing the (artificial) requirement that all the operators should be written in the normal form while the in the approach similar to Dirac's hole theory $\mathcal{E}_{vac}$ is necessarily an infinite negative constant. However, in GFQT the operators a and b are not independent and hence one cannot simply postulate that $\mathcal{E}_{vac} = 0$.

Consider first the sum in Eq. (8.81) when the values of n and k are fixed. It is convenient to distinguish the cases $s > 2k$ and $s < 2k$. If $s > 2k$ then, as follows from Eq. (8.28), the maximum value of n_1 is such that $m_{AdS} + 2(n + n_1)$ is always less than $2p$. For this reason all the values of n_1 contribute to the sum, which can be written as

$$\begin{aligned} S_1(n,k) = -\textstyle\sum_{n_1=0}^{p-q_1-n+k}[(m_{AdS} + 2n + 2n_1) + \\ (m_{AdS} + 2n + 2n_1 + 2) + ... + (2p-1)] \end{aligned} \tag{8.83}$$

A simple calculation shows that the result can be represented as

$$S_1(n,k) = \sum_{n_1=1}^{p-1} n_1^2 - \sum_{n_1=1}^{n+(m_{AdS}-3)/2} n_1^2 - \sum_{n_1=1}^{(s-1)/2-k} n_1^2 \tag{8.84}$$

where the last sum should be taken into account only if $(s-1)/2 - k \geq 1$.

The first sum in this expression equals $(p-1)p(2p-1)/6$ and, since we assume that $p \neq 2$ and $p \neq 3$, this quantity is zero in F_p. As a result, $S_1(n,k)$ is represented as a sum of two terms such that the first one depends only on n and the second — only on k. Note also that the second term is absent if $s = 1$, i.e. for particles with the spin 1/2 in the usual units.

Analogously, if $s < 2k$ the result is

$$S_2(n,k) = -\sum_{n_2=1}^{n+(m_{AdS}-3)/2} n_2^2 - \sum_{n_2=1}^{k-(s+1)/2} n_2^2 \tag{8.85}$$

where the second term should be taken into account only if $k-(s+1)/2 \geq 1$.

We now should calculate the sum

$$S(n) = \sum_{k=0}^{(s-1)/2} S_1(n,k) + \sum_{k=(s+1)/2}^{s} S_2(n,k) \tag{8.86}$$

and the result is

$$S(n) = -(s+1)(n+\frac{m_{AdS}-1}{2})[2(n+\frac{m_{AdS}-1}{2})^2 - \\ 3(n+\frac{m_{AdS}-1}{2})+1]/6 - (s-1)(s+1)^2(s+3)/96 \tag{8.87}$$

Since the value of n is in the range $[0, n_{max}]$, the final result is

$$E_{vac} = \sum_{n=0}^{n_{max}} S(n) = (m_{AdS}-3)(s-1)(s+1)^2(s+3)/96 \tag{8.88}$$

since in the massive case $n_{max} = p+2-m_{AdS}$.

Our final conclusion in this section is that *if s is odd and the separation of states into physical and nonphysical ones is accomplished as in Sec. 8.6 then $E_{vac} = 0$ only if $s = 1$ (i.e. $s = 1/2$ in the usual units).* This result shows that since the rules of arithmetic in Galois fields are different from that for real numbers, it is possible that quantities which are infinite in standard theory (e.g. the vacuum energy) will be zero in GFQT.

8.9 Neutral particles and spin-statistics theorem

In this section we will discuss the relation between the (a, a^*) and (b, b^*) operators only for all quantum numbers (i.e. in the spirit of the AB-symmetry) and therefore the results are valid regardless of whether the separation of states into S_+ and S_- can be justified or not (see the discussion in Sec. 8.7). In other words, we treat the set of the (b, b^*) operators not necessarily as the one related to antiparticles but simply as a set obtained from the (a, a^*) operators by the transformation defined by Eqs. (8.64) and (8.65).

The nonexistence of neutral elementary particles in GFQT is one of the most striking differences between GFQT and standard theory. One could give the following definition of neutral particle:

- i) it is a particle coinciding with its antiparticle

- ii) it is a particle which does not coincide with its antiparticle but they have the same properties

In standard theory only i) is meaningful since neutral particles are described by real (not complex) fields and this condition is required by Hermiticity. One might think that the definition ii) is only academic since if a particle and its antiparticle have the same properties then they are indistinguishable and can be treated as the same. However, the cases i) and ii) are essentially different from the operator point of view. In the case i) only the (a, a^*) operators are sufficient for describing the operators (8.59) in standard theory. This is the reflection of the fact that the real field has the number of degrees of freedom twice as less as the complex field. On the other hand, in the case ii) both (a, a^*) and (b, b^*) operators are required, i.e. in standard theory such a situation is described by a complex field. Nevertheless, the case ii) seems to be rather odd: it implies that there exists a quantum number distinguishing a particle from its antiparticle but this number is not manifested experimentally. We now consider whether the conditions i) or ii) can be implemented in GFQT.

Since each operator a is proportional to some operator b^* and vice versa (see Eqs. (8.64) and (8.65)), it is clear that if the particles described by the operators (a, a^*) have a nonzero charge then the particles described by the operators (b, b^*) have the opposite charge and the number of operators cannot be reduced. However, if all possible charges are zero, one could try to implement i) by requiring that each $b(n_1 n_2 nk)$ should be proportional to $a(n_1 n_2 nk)$ and then $a(n_1 n_2 nk)$ will be proportional to $a(\tilde{n}_1, \tilde{n}_2, nk)^*$. In this case the operators (b, b^*) will not be needed at all.

Suppose, for example, that the operators (a, a^*) satisfy the commutation relations (8.50). In that case the operators $a(n_1 n_2 nk)$ and $a(n_1' n_2' n' k')$ should commute if the sets $(n_1 n_2 nk)$ and $(n_1' n_2' n' k')$ are not the same. In particular, one should have $[a(n_1 n_2 nk), a(\tilde{n}_1 \tilde{n}_2 nk)] = 0$ if either $n_1 \neq \tilde{n}_1$ or $n_2 \neq \tilde{n}_2$. On the other hand, if $a(\tilde{n}_1 \tilde{n}_2 nk)$ is proportional to $a(n_1 n_2 nk)^*$, it follows from Eq. (8.50) that the commutator cannot be zero. Analogously one can consider the case of anticommutators.

The fact that the number of operators cannot be reduced is also clear from the observation that the (a, a^*) or (b, b^*) operators describe an irreducible representation in which the number of states (by definition) cannot be reduced. Our conclusion is that in GFQT the definition of neutral particle according to i) is fully unacceptable.

Consider now whether it is possible to implement the definition ii) in GFQT. Recall that we started from the operators (a, a^*) and defined the operators (b, b^*) by means of Eq. (8.64). Then the latter satisfy the same commutation or anticommutation relations as the former and the AB symmetry is valid. Does it mean that the particles described by the operators (b, b^*) are the same as the ones described by the operators (a, a^*)? If one starts from the operators (b, b^*) then, by analogy with Eq. (8.64), the operators (a, a^*) can be defined as

$$b(n_1 n_2 nk)^* = \eta'(n_1 n_2 nk) a(\tilde{n}_1 \tilde{n}_2 nk) / F(\tilde{n}_1 \tilde{n}_2 nk) \tag{8.89}$$

where $\eta'(n_1n_2nk)$ is some function. By analogy with the consideration in Sec. 8.5 one can show that

$$\eta'(n_1n_2nk) = \beta(-1)^{n_1+n_2+n}, \quad \beta\bar{\beta} = \mp 1 \tag{8.90}$$

where the minus sign refers to the normal spin-statistics connection and the plus to the broken one.

As follows from Eqs. (8.64), (8.67-8.70), (8.89), (8.90) and the definition of the quantities $\tilde{n}_1$ and $\tilde{n}_2$ in Sec. 8.5, the relation between the quantities α and β is $\alpha\bar{\beta} = 1$. Therefore, as follows from Eq. (8.90), there exist only two possibilities, $\beta = \mp\alpha$, depending on whether the normal spin-statistics connection is valid or not. We conclude that the broken spin-statistics connection implies that $\alpha\bar{\alpha} = \beta\bar{\beta} = 1$ and $\beta = \alpha$ while the normal spin-statistics connection implies that $\alpha\bar{\alpha} = \beta\bar{\beta} = -1$ and $\beta = -\alpha$. Since in the first case there exist solutions such that $\alpha = \beta$ (e.g. $\alpha = \beta = 1$), the particle and its antiparticle can be treated as neutral in the sense of the definition ii). Since such a situation is clearly unphysical, one might treat the Pauli spin-statistics theorem [11] as a requirement excluding neutral particles in the sense ii).

We now consider another possible treatment of the spin-statistics theorem, which seems to be much more interesting. In the case of the normal spin-statistics connection α satisfies Eq. (7.6). Such a relation is obviously impossible in standard theory.

As noted in Chap. 6, -1 is a quadratic residue in F_p if $p = 1 \ (mod\ 4)$ and a quadratic non-residue in F_p if $p = 3 \ (mod\ 4)$. For example, -1 is a quadratic residue in F_5 since $2^2 = -1 \ (mod\ 5)$ but in F_7 there is no element a such that $a^2 = -1 \ (mod\ 7)$. We conclude that if $p = 1 \ (mod\ 4)$ then Eq. (7.6) has solutions in F_p and in that case the theory can be constructed without any extension of F_p.

Consider now the case $p = 3 \ (mod\ 4)$. Then Eq. (7.6) has no solutions in F_p and it is necessary to consider this equation in an extension of F_p (*i.e.*, there is no "real" version of GFQT). The minimum extension is obviously F_{p^2} and therefore the problem arises whether Eq. (7.6) has solutions in F_{p^2}. As shown in Sec. 7.1, this equation does have solutions.

Our conclusion is that *if* $p = 3 \ (mod\ 4)$ *then the spin-statistics theorem implies that the field* F_p *should necessarily be extended and the minimum possible extension is* F_{p^2}. Therefore the spin-statistics theorem can be treated as a requirement that GFQT should be based on F_{p^2} and standard theory should be based on complex numbers.

Let us now discuss a different approach to the AB symmetry. A desire to have operators which can be interpreted as those relating separately to particles and antiparticles is natural in view of our experience in standard approach. However, one might think that in the spirit of GFQT there is no need to have separate operators for particles and antiparticles since they are different states of the same object. We can therefore reformulate the AB symmetry in terms of only (a, a^*) operators as follows.

Instead of Eqs. (8.64) and (8.65), we consider a *transformation* defined as

$$a(n_1n_2nk)^* \to \eta(n_1n_2nk)a(\tilde{n}_1\tilde{n}_2nk)/F(\tilde{n}_1\tilde{n}_2nk)$$
$$a(n_1n_2nk) \to \bar{\eta}(n_1n_2nk)a(\tilde{n}_1\tilde{n}_2nk)^*/F(\tilde{n}_1\tilde{n}_2nk) \quad (8.91)$$

Then the AB symmetry can be formulated as a requirement that physical results should be invariant under this transformation.

Let us now apply the AB transformation twice. Then we get

$$a(n_1n_2nk)^* \to \mp a(n_1n_2nk)^*, \quad a(n_1n_2nk) \to \mp a(n_1n_2nk) \quad (8.92)$$

for the normal and broken spin-statistic connections, respectively. Therefore, as a consequence of the spin-statistics theorem, any particle (with the integer or half-integer spin) has the AB^2 parity equal to -1. Therefore in GFQT any interaction can involve only an even number of creation and annihilation operators. In particular, this is additional demonstration of the fact that in GFQT the existence of neutral elementary particles is incompatible with the spin-statistics theorem.

8.10 Modular IRs of the osp(1,4) superalgebra

If one accepts supersymmetry then the results on modular IRs of the so(2,3) algebra can be generalized by considering modular IRs of the osp(1,4) superalgebra. Representations of the osp(1,4) superalgebra have several interesting distinctions from representations of the Poincare superalgebra. For this reason we first briefly mention some known facts about the latter representations (see e.g Ref. [110] for details).

Representations of the Poincare superalgebra are described by 14 operators. Ten of them are the representation operators of the Poincare algebra—four momentum operators and six representation operators of the Lorentz algebra, which satisfy the commutation relations (1.3). In addition, there are four fermionic operators. The anticommutators of the fermionic operators are linear combinations of the momentum operators, and the commutators of the fermionic operators with the Lorentz algebra operators are linear combinations of the fermionic operators. In addition, the fermionic operators commute with the momentum operators.

From the formal point of view, representations of the osp(1,4) superalgebra are also described by 14 operators — ten representation operators of the so(2,3) algebra and four fermionic operators. There are three types of relations: the operators of the so(2,3) algebra commute with each other as usual (see Sec. 8.2), anticommutators of the fermionic operators are linear combinations of the so(2,3) operators and commutators of the latter with the fermionic operators are their linear combinations. However, in fact representations of the osp(1,4) superalgebra can be described exclusively in terms of the fermionic operators. The matter is as follows. In the general case the anticommutators of four operators form ten independent linear combinations. Therefore, ten bosonic operators can be expressed in terms of fermionic ones. This is not the case for the Poincare superalgebra since the Poincare algebra operators are obtained from the so(2,3) one by contraction. One can say that the representations of the osp(1,4) superalgebra is an implementation of the idea that supersymmetry is the extraction of the square root from the usual symmetry (by analogy with the treatment of the Dirac equation as a square root from the Klein-Gordon one).

We use $(d'_1, d'_2, d_1", d_2")$ to denote the fermionic operators of the osp(1,4) superalgebra. They should satisfy the following relations. If (A, B, C) are any fermionic operators, [...,...] is used to denote a commutator and $\{..., ...\}$ to denote an anticommutator then

$$[A, \{B, C\}] = F(A, B)C + F(A, C)B \tag{8.93}$$

where the form $F(A, B)$ is skew symmetric, $F(d'_j, d_j") = 1$ $(j = 1, 2)$ and the other independent values of $F(A, B)$ are equal to zero. The fact that the representation of the osp(1,4) superalgebra is fully defined by Eq. (8.93) and the properties of the form $F(.,.)$, shows that osp(1,4) is a special case of the superalgebra.

We can now **define** the so(2,3) operators as follows:

$$\begin{gathered} b' = \{d'_1, d'_2\}, \quad b" = \{d_1", d_2"\}, \quad L_+ = \{d'_2, d_1"\}, \quad L_- = \{d'_1, d_2"\} \\ a'_j = (d'_j)^2, \quad a_j" = (d_j")^2, \quad h_j = \{d'_j, d_j"\} \quad (j = 1, 2) \end{gathered} \tag{8.94}$$

Then by using Eq. (8.93) and the properties of the form $F(.,.)$, one can show by direct calculations that so defined operators satisfy the commutation relations (8.9,8.13,8.14). This result can be treated as a fact that the operators of the so(2,3) algebra are not fundamental, only the fermionic operators are.

By analogy with the construction of IRs of the osp(1,4) superalgebra in standard theory [111], we require the existence of the generating vector e_0 satisfying the conditions :

$$d'_j e_0 = d'_2 d_1" e_0 = 0, \quad d'_j d_j" e_0 = q_j e_0 \quad (j = 1, 2) \tag{8.95}$$

These conditions are written exclusively in terms of the d operators. As follows from Eq. (8.94), they can be rewritten as (compare with Eq. (8.19))

$$d'_j e_0 = L_+ e_0 = 0, \quad h_j e_0 = q_j e_0 \quad (j = 1, 2) \tag{8.96}$$

The full representation space can be obtained by successively acting by the fermionic operators on e_0 and taking all possible linear combinations of such vectors.

We use E to denote an arbitrary linear combination of the vectors $(e_0, d_1" e_0, d_2" e_0, d_2" d_1" e_0)$. Our next goal is to prove a statement analogous to that in Ref. [111]:

Statement 1: Any vector from the representation space can be represented as a linear combination of the elements $O_1 O_2 ... O_n E$ where $n = 0, 1, ...$ and O_i is an operator of the so(2,3) algebra.

The first step is to prove a simple

Lemma: If D is any fermionic operator then DE is a linear combination of elements E and OE where O is an operator of the so(2,3) algebra.

The proof is by a straightforward check using Eqs. (8.93-8.96). For example,

$$d_1"(d_2"d_1"e_0) = \{d_1", d_2"\}d_1"e_0 - d_2"a_1"e_0 = b"d_1"e_0 - a_1"d_2"e_0$$

To prove Statement 1 we define the height of a linear combination of the elements $O_1O_2...O_nE$ as the maximum sum of powers of the fermionic operator in this element. For example, since each operator of the so(2,3) algebra is composed of two fermionic operator, the height of the element $O_1O_2...O_nE$ equals $2n+2$ if E contains $d_2"d_1"e_0$, equals $2n+1$ if E does not contain $d_2"d_1"e_0$ but contains either $d_1"e_0$ or $d_2"e_0$ and equals $2n$ if E contains only e_0.

We can now prove Statement 1 by induction. The elements with the heights 0, 1 and 2 obviously have the required form since, as follows from Eq. (8.94), $d_1"d_2"e_0 = b"e_0 - d_2"d_1"e_0$. Let us assume that Statement 1 is correct for all elements with the heights $\leq N$. Every element with the height $N+1$ can be represented as Dx where x is an element with the height N. If $x = O_1O_2...O_nE$ then by using Eq. (8.93) we can represent Dx as $Dx = O_1O_2...O_nDE + y$ where the height of the element y is $N-1$. As follows from the induction assumption, y has the required form, and, as follows from Lemma, DE is a linear combination of the elements E and OE. Therefore Statement 1 is proved.

As follows from Eqs. (8.93) and (8.94),

$$[d'_j, h_j] = d'_j, \quad [d_j", h_j] = -d_j", \quad [d'_j, h_l] = [d_j", h_l] = 0 \quad (j,l = 1,2 \; j \neq l) \quad (8.97)$$

It follows from these expressions that if x is such that $h_jx = \alpha_jx$ $(j = 1,2)$ then $d_1"x$ is the eigenvector of the operators h_j with the eigenvalues (α_1+1, α_2), $d_2"x$ - with the eigenvalues (α_1, α_2+1), d'_1x - with the eigenvalues (α_1-1, α_2), and d'_2x - with the eigenvalues α_1, α_2-1.

By analogy with the case of IRs of the so(2,3) algebra (see Sec. 8.2), we assume that q_1 and q_2 are represented by the numbers $0, 1, ...p-1$. We first consider the case when $q_2 \geq 1$ and $q_1 \geq q_2$. We again use m_{AdS} to denote q_1+q_2 and s to denote q_1-q_2. We first assume that $m_{AdS} \neq 2$ and $s \neq p-1$. Then Statement 1 obviously remains valid if we now assume that E contains linear combinations of (e_0, e_1, e_2, e_3) where

$$e_1 = d_1"e_0, \quad e_2 = [d_2" - \frac{1}{s+1}L_-d_1"]e_0$$
$$e_3 = (d_2"d_1"e_0 - \frac{q_1-1}{m_{AdS}-2}b" + \frac{1}{m_{AdS}-2}a_1"L_-)e_0 \quad (8.98)$$

As follows from Eqs. (8.93-8.97), e_0 satisfies Eq. (8.19) and e_1 satisfies the same condition with q_1 replaced by q_1+1. We see that the representation of the osp(1,4) superalgebra defined by Eq. (8.96) necessarily contains at least two IRs of the so(2,3) algebra characterized by the values of the mass and spin (m_{AdS}, s) and $(m_{AdS}+1, s+1)$ and the generating vectors e_0 and e_1, respectively.

As follows from Eqs. (8.93-8.97), the vectors e_2 and e_3 satisfy the conditions

$$h_1e_2 = q_1e_2, \quad h_2e_2 = (q_2+1)e_2, \quad h_1e_3 = (q_1+1)e_3, \quad h_2e_3 = (q_2+1)e_3$$
$$a_1'e_j = a_2'e_j = b'e_j = L_+e_j = 0 \quad (j=2,3) \tag{8.99}$$

and therefore (see Eq. (8.19)) they will be generating vectors of IRs of the so(2,3) algebra if they are not equal to zero.

If $s=0$ then, as follows from Eqs. (8.93,8.94,8.98), $e_2=0$. In the general case, as follows from these expressions,

$$d_1'e_2 = \frac{1-q_2}{s+1}L_-e_0, \quad d_2'e_2 = \frac{s(q_2-1)}{s+1}e_0 \tag{8.100}$$

Therefore e_2 is also a null vector if e_0 belongs to the massless IR (with $q_2 = 1$) while $e_2 \neq 0$ if $s \neq 0$ and $q_2 \neq 1$. As follows from direct calculation using Eqs. (8.93,8.94,8.98)

$$d_1'e_3 = \frac{m_{AdS}-1}{m_{AdS}-2}[L_-d_1" - (2q_2+s-1)d_2"]e_0, \quad d_2'e_3 = (q_2 - \frac{q_1-1}{m_{AdS}-2})e_0 \tag{8.101}$$

If $q_2 = 1$ then $d_1'e_3$ is proportional to e_2 (see Eq. (8.98)) and hence $d_1'e_3 = 0$. In this case $q_1 - 1 = m_{AdS} - 2$ and hence $d_2'e_3 = 0$. Therefore we conclude that $e_3 = 0$. It is also clear from Eq. (8.101) that $e_3 = 0$ if $m_{AdS} = 1$. In all other cases $e_3 \neq 0$.

Consider now the case $m_{AdS} = 2$. If $s=0$ then $q_1 = q_2 = 1$. The condition $e_2 = 0$ is still valid for the same reasons as above but if e_3 is defined as $[d_2", d_1"]e_0/2$ then e_3 is the minimal $sp(2)\times sp(2)$ vector with $h_1 = h_2 = 2$ and, as a result of direct calculations using Eqs. (8.93,8.94,8.98)

$$d_1'e_3 = \frac{1}{2}(1-2q_1)d_2"e_0, \quad d_2'e_3 = \frac{1}{2}(2q_2-1)e_0 \tag{8.102}$$

Hence in this case $e_3 \neq 0$ and the IR of the osp(1,4) superalgebra corresponding to $(q_1, q_2) = (1,1)$ contains IRs of the so(2,3) algebra corresponding to $(1,1)$, $(2,1)$ and $(2,2)$. Therefore this IR of the osp(1,4) superalgebra should be treated as massive rather than massless.

At this point the condition that q_1 and q_2 are taken modulo p has not been explicitly used and, as already mentioned, our considerations are similar to those in Ref. [111]. Therefore when $q_1 \geq q_2$, modular IRs of the osp(1,4) superalgebra can be characterized in the same way as conventional IRs [111, 112]:

- If $q_2 > 1$ and $s \neq 0$ (massive IRs), the osp(1,4) supermultiplets contain four IRs of the so(2,3) algebra characterized by the values of the mass and spin $(m,s), (m+1,s+1), (m+1,s-1), (m+2,s)$.

- If $q_2 \geq 1$ and $s = 0$ (collapsed massive IRs), the osp(1,4) supermultiplets contain three IRs of the so(2,3) algebra characterized by the values of the mass and spin $(m, s), (m+1, s+1), (m+2, s)$.
- If $q_2 = 1$ and $s = 1, 2, ...p-2$ (massless IRs) the osp(1,4) supermultiplets contains two IRs of the so(2,3) algebra characterized by the values of the mass and spin $(2+s, s), (3+s, s+1)$.
- Dirac supermultiplet containing two Dirac singletons (see Sec. 8.3).

The first three cases have well-known analogs of IRs of the super-Poincare algebra (see e.g., Ref. [110]) while there is no super-Poincare analog of the Dirac supermultiplet.

Since the space of IR of the superalgebra osp(1,4) is a direct sum of spaces of IRs of the so(2,3) algebra, for modular IRs of the osp(1,4) superalgebra one can prove results analogous to those discussed in the preceding sections. In particular, one modular IR of the osp(1,4) algebra is a modular analog of both standard IRs of the osp(1,4) superalgebra with positive and negative energies. This implies that one modular IR of the osp(1,4) superalgebra contains both, a superparticle and its anti-superparticle.

At the same time, as noted in Sec. 8.2, there are special cases which have no analogs in standard theory. The above results can be applied to those cases without any changes. For example, the special singleton characterized by $(m_{AdS} = 0, s)$, $s \neq 0$ generates a special supersingleton containing IRs of the so(2,3) algebra with $(m_{AdS} = 0, s)$, $(m_{AdS} = 1, s+1)$, $(m_{AdS} = 1, s-1)$ and $(m_{AdS} = 2, s)$. In particular, when $s = 1$ then two of those IRs are the Di and Rac singletons. All other special singletons also generate supersingletons containing more than two IRs of the so(2,3) algebra. Hence the Dirac supersingleton can be treated as a more fundamental object than other special supersingletons. For this reason, among supersingletons we will consider only the case of the Dirac supersingleton. Then we will see below that the decomposition of the tensor product of the Dirac supersingletons can contain only special IRs of the osp(1,4) superalgebra with $q_1 = 0$. In this case we have that $d_1' d_1'' e_0 = q_1 e_0 = 0$, $d_2' d_1'' e_0 = L_+ e_0 = 0$ and hence $d_1'' e_0 = 0$. Since $L_+ d_2'' e_0 = d_1'' e_0 = 0$ and $d_2' d_2'' e_0 = q_2 e_0$, the vector $d_2'' e_0$ is not zero and if e_0 is the generating vector for the IR of the so(2,3) algebra with $(q_1 = 0, q_2)$ then $d_2'' e_0$ is the generating vector for the IR of the so(2,3) algebra with $(0, q_2 + 1)$. The IR of the osp(1,4) superalgebra does not contain other IRs of the so(2,3) algebra since $d_2'' d_1'' e_0 = 0$ and $d_1'' d_2'' e_0 = (d_1'' d_2'' + d_2'' d_1'') e_0 = b'' e_0$.

By analogy with Sec. 8.3, we use $SDim(s)$ to denote the dimension of the IR of the osp(1,4) superalgebra in the massless case with the spin s and $SDim(q_1, q_2)$ to denote the dimension of the IR of the osp(1,4) superalgebra characterized by the quantities q_1 and q_2. Then as follows from the above discussion

$$\begin{aligned}
&SDim(0,q_2) = Dim(0,q_2) + Dim(0,q_2+1) \quad (q_2 = 1,2,...p-1) \\
&SDim(s) = Dim(s) + Dim(s+1) \quad (s = 1,2,...p-2) \\
&SDim(1,1) = Dim(1,1) + Dim(2,1) + Dim(2,2)
\end{aligned} \tag{8.103}$$

and $Dim(p-1) = Dim(0,1)$.

Chapter 9

Dirac singletons as the only true elementary particles

9.1 Why Dirac singletons are indeed remarkable

As already noted, Dirac singletons have been discovered by Dirac in his paper [109] titled "A remarkable representation of the 3 + 2 de Sitter group". In this section we argue that in GFQT the Dirac singletons are even more remarkable than in standard theory. As noted in Sec. 8.2, in the theory over a Galois field there also exist special singleton-like IRs which have no analogs in standard theory. As argued in Sec. 8.10, from the point of view of supersymmetry they are less fundamental than Dirac singletons. For this reason we will not consider such IRs and the term singleton will always mean the Dirac singleton.

Although theory of elementary particles exists for a rather long period of time, it has been noted in Sec. 3.2 that there is no commonly accepted definition of elementary particle in this theory. In the preceding chapters we adopted the definition that a particle is elementary if it is described by an IR of the symmetry algebra (in standard theory this IR is implemented by Hermitian operators while in GFQT it is a representation over a Galois field).

As shown in Sec. 8.2, each IR of the so(2,3) algebra is characterized by the quantities (q_1, q_2). Consider a system of two particles such that the IR describing particle 1 is defined by the numbers $(q_1^{(1)}, q_2^{(1)})$ and the IR describing particle 2 is defined by the numbers $(q_1^{(2)}, q_2^{(2)})$. The representation describing such a system is the tensor product of the corresponding IRs defined as follows. Let $\{e_i^{(1)}\}$ and $\{e_j^{(2)}\}$ be the sets of basis vectors for the IRs describing particle 1 and 2, respectively. Then the basis of the tensor product is formed by the elements $e_{ij} = e_i^{(1)} \times e_j^{(2)}$. Let $\{O_k^{(1)}\}$ and $\{O_l^{(2)}\}$ $(k, l = 1, 2, ...10)$ be the sets of independent representation operators in the corresponding IRs. Then the set of independent representation operators in the tensor product is $\{O_k = O_k^{(1)} + O_k^{(2)}\}$. Here it is assumed that the operator with the superscript (j) acts on the elements $e_k^{(j)}$ in the same way as in the IR j while on the elements $e_l^{(j')}$ where $j' \neq j$ it acts as the identity operator. For example,

$$h_1 \sum_{ij} c_{ij}(e_i^{(1)} \times e_j^{(2)}) = \sum_{ij} c_{ij}[(h_1^{(1)} e_i^{(1)}) \times e_j^{(2)} + e_i^{(1)} \times (h_1^{(2)} e_j^{(2)})]$$

Then the operators $\{O_k\}$ satisfy the same commutation relations as in Eqs. (8.9), (8.13) and (8.14).

It is immediately clear from this definition that the tensor product of IRs characterized by $(q_1^{(1)}, q_2^{(1)})$ and $(q_1^{(2)}, q_2^{(2)})$, respectively, contains at least the IR characterized by $(q_1 = q_1^{(1)} + q_1^{(2)}, q_2 = q_2^{(1)} + q_2^{(2)})$. Indeed, if $e_0^{(j)}$ $(j = 1, 2)$ is the generating vector for IR j then the vector $e_0 = e_0^{(1)} \times e_0^{(2)}$ will satisfy Eq. (8.19). In turn, states of an elementary particle characterized by (q_1, q_2) can be constructed as composite states of two elementary particles characterized by $(q_1^{(1)}, q_2^{(1)})$ and $(q_1^{(2)}, q_2^{(2)})$, respectively, if $(q_1 = q_1^{(1)} + q_1^{(2)}, q_2 = q_2^{(1)} + q_2^{(2)})$.

This poses a question whether the notions of elementary and composite particles are absolute. For example, as noted in Sec. 8.2, massless particles are characterized such that $q_2 = 1$ and in GFQT the quantity q_2 characterizing a massive particle is such that $(q_2 = 2, 3, ...p-1)$. Hence a massive particle characterized by $q_2 = n$ can be constructed as a composite state of n massless particles. In Standard Model (based on Poincare invariance) only massless particles are treated as elementary. However, as shown in a seminal paper by Flato and Fronsdal [113] (see also Ref. [114]), in standard AdS theory each massless IR can be constructed as a tensor product of two singleton IRs and the authors of Ref. [113] believe that this a truly remarkable property. On the other hand, since the Rac IR is characterized by $q_1 = q_2 = 1/2$, it can be constructed as a composite state of two massive IRs characterized by $q_1 = q_2 = 1/4$ where in standard theory $1/4$ is understood as a rational number and in GFQT - as an element of F_p. Analogously the Di IR can be constructed as a composite state of two IRs characterized by $(q_1 = 3/4, q_2 = 1/4)$.

In general, it is obvious that in standard theory an IR characterized by (q_1, q_2) can be constructed from tensor products of two IRs characterized by $(q_1^{(1)}, q_2^{(1)})$ and $(q_1^{(2)}, q_2^{(2)})$ only if $q_1 \geq (q_1^{(1)} + q_1^{(2)})$ and $q_2 \geq (q_2^{(1)} + q_2^{(2)})$. Since no interaction is assumed, a problem arises whether a particle constructed from a tensor product of other two particles will be stable. In standard theory a particle with the mass m can be a stable composite state of two particles with the masses m_1 and m_2 only if $m < (m_1 + m_2)$ and the quantity $(m_1 + m_2 - m)c^2$ is called the binding energy. The greater the binding energy is the more stable is the composite state with respect to external interactions.

In view of the above discussion, a particle can be called elementary if it is not a composite state of other particles in the given theory. If a theory is formulated in terms of a Fock space, this implies that only those particles are treated as elementary whose operators (a, a^*) are used in the description of the Fock space.

The authors of Ref. [113] and other works treat singletons as true elementary particles because their weight diagrams has only a single trajectory (that's why the corresponding IRs are called singletons) and in AdS QFT singleton fields live on the boundary at infinity of the AdS bulk (boundary which has one dimension less than the bulk). However, in that case one should should answer the following questions:

- a) Why singletons have not been observed yet.
- b) Why such massless particles as photons and others are stable and their decays into singletons have not been observed.

There exists a wide literature (see e.g. Ref. [115, 116] and references therein) where this problem is investigated from the point of view of standard AdS QFT. However, as noted in Sec. 1.2, the physical meaning of field operators is not clear and products of local quantized fields at the same points are not well defined.

In addition, the following question arises. Each massless boson (e.g. the photon) can be constructed as a tensor product of either two Dis or two Racs. Which of those possibilities (if any) is physically preferable? A natural answer is as follows. If the theory is supersymmetric then the AdS algebra should be extended to the superalgebra osp(1,4) which has only one positive energy IR combining Di and Rac into the Dirac supermultiplet. For the first time, this possibility has been discussed probably in Refs. [112, 111]. Therefore in standard theory there exists only one Dirac superparticle and its antiparticle.

As shown in the preceding chapter, in GFQT one IRs describes a particle and its antiparticle simultaneously and hence in GFQT there exists only one IR describing the supersingleton. In addition, as shown in Sec. 8.3, while dimensions of massless IRs are of the order of p^3 (see Eqs. (8.37-8.39)), the dimensions of the singleton IRs are of the order of p^2 (see Eq. (8.35)) and, as follows from Eq. (8.35), the dimension of the supersingleton IR is p^2. These facts can be treated as arguments that in GFQT the supersingleton can be the only elementary particle. In addition, in Chap. 10 we argue that, in contrast to standard theory, in GFQT one can give natural explanations of a) and b).

The chapter is organized as follows. In Sec. 9.3 we discuss in detail how usual particles and singletons should be discussed in the Poincare and semiclassical limits of standard theory. In Sec. 9.4 it is shown that, in contrast to standard theory, the tensor products of singleton IRs in GFQT contain not only massless IRs but also special IRs, which have no analogs in standard theory. Beginning from Sec. 9.5 we proceed to the supersymmetric case, and the main result of the chapter is described in Sec. 9.6. Here we explicitly find a complete list of IRs taking part in the decomposition of the tensor product of two supersingletons. In standard theory the well-known results are recovered while in GFQT this list also contains special supersymmetric IRs which have no analogs in standard theory.

9.2 Tensor product of modular IRs of the sp(2) algebra

Consider two modular IRs of the sp(2) algebra in spaces H_j ($j = 1,2$). Each IR is defined by a set of operators $(h^{(j)}, a^{(j)\prime}, a^{(j)"})$ satisfying the commutation relations (8.1) and by a vector $e_0^{(j)}$ such that (see Eq. (8.3))

$$a^{(j)\prime}e_0^{(j)} = 0, \quad h^{(j)}e_0 = q_0^{(j)}e_0^{(j)} \tag{9.1}$$

As follows from the results of the preceding section, the vectors $e_n^{(j)} = (a^{(j)"})^n e_0^{(j)}$ where $k = 0,1,...N^{(j)}$ and $N^{(j)} = p - q_0^{(j)}$ form a basis in H_j.

The tensor product of such IRs can be defined by analogy with the definition of the tensor product of IRs of the so(2,3) algebra in the preceding section. The basis of the representation space is formed by the elements $e_{kl} = e_k^{(1)} \times e_l^{(2)}$ and the independent representation operators are $(h, a', a")$ such that $h = h^{(1)} + h^{(2)}$, $a' = a^{(1)\prime} + a^{(2)\prime}$ and $a" = a^{(1)"} + a^{(2)"}$. Then the operators $(h, a', a")$ satisfy the same commutation relations as in Eq. (8.1) and hence they implement a representation of the sp(2) algebra in the space $H_1 \times H_2$. Our goal is to find a decomposition of this representation into irreducible components.

It is obvious that the cases when $q_0^{(1)} = 0$ or $q_0^{(2)} = 0$ are trivial and therefore we will assume that $q_0^{(1)} \neq 0$ and $q_0^{(2)} \neq 0$. If $q_0^{(1)}$ and $q_0^{(2)}$ are represented by the numbers $(1,2,...p-1)$ then we suppose that $q_0^{(1)} \geq q_0^{(2)}$ and consider the vector

$$e(k) = \sum_{i=0}^{k} c(i,k)(e_i^{(1)} \times e_{k-i}^{(2)}) \tag{9.2}$$

As follows from Eq. (8.4) and the definition of h,

$$he(k) = (q_0^{(1)} + q_0^{(2)} + 2k)e(k) \tag{9.3}$$

Therefore if $a'e(k) = 0$ then the vector $e(k)$ generates a modular IR with the dimension $Dim(q_0^{(1)}, q_0^{(2)}, k) = p + 1 - (q_0^{(1)} - q_0^{(2)} - 2k)$ where $q_0^{(1)} - q_0^{(2)} - 2k$ is taken modulo p. As follows from Eqs. (8.5) and (9.2),

$$a'e(k) = \sum_{i=0}^{k} c(i,k)[i(q_0^{(1)}+i-1)(e_{i-1}^{(1)} \times e_{k-i}^{(2)}) + (k-i)(q_0^{(2)}+k-i-1)(e_i^{(1)} \times e_{k-i-1}^{(2)})] \tag{9.4}$$

This condition will be satisfied if

$$c(i,k) = \frac{(k+1-i)(q_0^{(2)}+k-i)c(i-1,k)}{i(q_0^{(1)}+i-1)} \quad (i = 1,...k) \tag{9.5}$$

It is clear from this expression that in standard case the possible values of k are $0,1,...\infty$ while in modular case $k=0,1,...k_{max}$ where $k_{max}=p-q_0^{(1)}$.

It is obvious that at different values of k, the IRs generated by $e(k)$ are linearly independent and therefore the tensor product of the IRs generated by $e_0^{(1)}$ and $e_0^{(2)}$ contains all the IRs generated by $e(k)$. A question arises whether the latter IRs give a full decomposition of the tensor product. This is the case when the dimension of the tensor product equals the sum of dimensions of the IRs generated by $e(k)$. Below we will be interested in the tensor product of singleton IRs and, as shown in Sec. 8.3, in that case $q_0^{(1)}+q_0^{(2)}>p$. Therefore $q_0^{(1)}+q_0^{(2)}+2k\in[q_0^{(1)}+q_0^{(2)},2p-q_0^{(1)}+q_0^{(2)}]$ and for all values of k, $q_0^{(1)}+q_0^{(2)}+2k$ is in the range $(p,2p]$. Then, as follows from Eq. (8.7), the fact that the IRs generated by $e(k)$ give a full decomposition of the tensor product follows from the relation

$$\sum_{k=0}^{p-q_0^{(1)}}(2p+1-q_0^{(1)}-q_0^{(2)}-2k)=(p+1-q_0^{(1)})(p+1-q_0^{(2)}) \tag{9.6}$$

9.3 Semiclassical approximation in Poincare limit

The Flato-Fronsdal result [113] poses a fundamental question whether only singletons can be true elementary particles. In the present work we consider singletons from the point of view of a quantum theory over a Galois field (GFQT) but the approach is applicable in standard theory (over the complex numbers) as well. As already noted in Sec. 8.3, the properties of singletons in standard theory and GFQT are considerably different. In this chapter and Chap. 10 we argue that in GFQT the singleton physics is even more interesting than in standard theory. However, since there exists a wide literature on singleton properties in standard theory, in the present section we discuss what conclusions can be made about semiclassical approximation and Poincare limit for singletons in this theory.

The first step is to obtain expressions for matrix elements of representation operators. Since spin is a pure quantum phenomenon, one might expect that in semiclassical approximation it suffices to consider the spinless case. Then, as shown in Sec. 8.2, the quantum number k can take only the value $k=0$, the basis vectors of the IR can be chosen as $e(n_1n_2n)=(a_1")^{n_1}(a_2")^{n_2}e_n$ (compare with Eq. (8.27)) where (see Eq. (8.5)) $e_n=(A^{++})^ne_0$. In the spinless case, $q_1=q_2=m/2$ and hence Eqs. (8.42-8.48) can be rewritten in the form:

$$h_1e(n_1n_2n)=[Q+2n_1]e(n_1n_2n)\quad h_2e(n_1n_2n)=[Q+2n_2]e(n_1n_2n) \tag{9.7}$$

$$\begin{aligned} a_1'e(n_1n_2n)&=n_1[Q+n_1-1]e(n_1-1,n_2n)\ a_1"e(n_1n_2n)=e(n_1+1,n_2n)\\ a_2'e(n_1n_2n)&=n_2[Q+n_2-1]e(n_1,n_2-1,n)\ a_2"e(n_1n_2n)=e(n_1,n_2+1,n) \end{aligned} \tag{9.8}$$

$$b"e(n_1n_2n) = \frac{Q-2}{Q-1}n(m_{AdS}+n-3)e(n_1+1,n_2+1,n-1) + \frac{1}{(Q-1)^2}e(n_1,n_2,n+1) \quad (9.9)$$

$$b'e(n_1n_2n) = \frac{Q-2}{Q-1}n(m_{AdS}+n-3)(Q+n_1-1)(Q+n_2-1)e(n_1,n_2,n-1) + \frac{n_1n_2}{(Q-1)^2}e(n_1-1,n_2-1,n+1) \quad (9.10)$$

$$L_+e(n_1n_2n) = \frac{Q-2}{Q-1}n(m_{AdS}+n-3)(Q+n_2-1)e(n_1+1,n_2,n-1) + \frac{n_2}{(Q-1)^2}e(n_1,n_2-1,n+1) \quad (9.11)$$

$$L_-e(n_1n_2n) = \frac{Q-2}{Q-1}n(m_{AdS}+n-3)(Q+n_1-1)e(n_1,n_2+1,n-1) + \frac{n_1}{(Q-1)^2}e(n_1-1,n_2,n+1) \quad (9.12)$$

where $Q = Q(n) = m_{AdS}/2 + n$.

The basis elements $e(n_1n_2n)$ are not normalized to one and in our special case the results given by Eqs. (8.29-8.31) can be represented as

$$||e(n_1n_2n)|| = F(n_1n_2n) = \{n!(m_{AdS}-2)_n[(\frac{m_{AdS}}{2})_n]^3(\frac{m_{AdS}}{2}-1)_n \, n_1!n_2!(\frac{m_{AdS}}{2}+n)_{n_1}(\frac{m_{AdS}}{2}+n)_{n_2}\}^{1/2} \quad (9.13)$$

By using this expression, Eqs. (9.7-9.12) can be rewritten in terms of the matrix elements of representation operators with respect to the normalized basis $\tilde{e}(n_1n_2n) = e(n_1n_2n)/F(n_1n_2n)^{1/2}$.

Each element of the representation space can be written as

$$x = \sum_{n_1n_2n} c(n_1n_2n)\tilde{e}(n_1n_2n)$$

where $c(n_1n_2n)$ can be called the wave function in the (n_1n_2n) representation. It is normalized as

$$\sum_{n_1n_2n} |c(n_1n_2n)|^2 = 1$$

In standard theory the quantum numbers n_1 and n_2 are in the range $[0,\infty)$. For massive and massless particles the quantum number n also is in this range while, as shown in Sec. 8.2, the only possible values of n for the spinless Rac singleton are $n = 0, 1$. By using Eqs. (9.7-9.13), one can obtain the action of the representation operator on the wave function $c(n_1n_2n)$:

$$\begin{aligned}
h_1c(n_1n_2n) &= [m_{AdS}/2+n+2n_1]c(n_1n_2n)\\
h_2c(n_1n_2n) &= [m_{AdS}/2+n+2n_2]c(n_1n_2n)\\
a_1'c(n_1n_2n) &= [(n_1+1)(m_{AdS}/2+n+n_1)]^{1/2}c(n_1+1,n_2n)\\
a_1"c(n_1n_2n) &= [n_1(m_{AdS}/2+n+n_1-1)]^{1/2}c(n_1-1,n_2n)\\
a_2'c(n_1n_2n) &= [(n_2+1)(m_{AdS}/2+n+n_2)]^{1/2}c(n_1,n_2+1,n)\\
a_2"c(n_1n_2n) &= [n_2(m_{AdS}/2+n+n_2-1)]^{1/2}c(n_1,n_2-1,n)\\
b"c(n_1n_2n) &= [\tfrac{n(m_{AdS}+n-3)(m_{AdS}/2+n+n_1-1)(m_{AdS}/2+n+n_2-1)}{(m_{AdS}/2+n-1)(m_{AdS}/2+n-2)}]^{1/2}c(n_1,n_2,n-1)+\\
&[\tfrac{n_1n_2(n+1)(m_{AdS}+n-2)}{(m_{AdS}/2+n)(m_{AdS}/2+n-1)}]^{1/2}c(n_1-1,n_2-1,n+1)\\
b'c(n_1n_2n) &= [\tfrac{(n+1)(m_{AdS}+n-2)(m_{AdS}/2+n+n_1)(m_{AdS}/2+n+n_2)}{(m_{AdS}/2+n)(m_{AdS}/2+n-1)}]^{1/2}c(n_1,n_2,n+1)+\\
&[\tfrac{(n_1+1)(n_2+1)n(m_{AdS}+n-3)}{(m_{AdS}/2+n-1)(m_{AdS}/2+n-2)}]^{1/2}c(n_1+1,n_2+1,n-1)\\
L_+c(n_1n_2n) &= [\tfrac{(n+1)(m_{AdS}+n-2)n_1(m_{AdS}/2+n+n_2)}{(m_{AdS}/2+n)(m_{AdS}/2+n-1)}]^{1/2}c(n_1-1,n_2,n+1)+\\
&[\tfrac{(n_2+1)n(m_{AdS}+n-3)(m_{AdS}/2+n+n_1-1)}{(m_{AdS}/2+n-1)(m_{AdS}/2+n-2)}]^{1/2}c(n_1,n_2+1,n-1)\\
L_-c(n_1n_2n) &= [\tfrac{n(m_{AdS}+n-3)(n_1+1)(m_{AdS}/2+n+n_2-1)}{(m_{AdS}/2+n-1)(m_{AdS}/2+n-2)}]^{1/2}c(n_1+1,n_2,n-1)+\\
&[\tfrac{n_2(n+1)(m_{AdS}+n-2)(m_{AdS}/2+n+n_1)}{(m_{AdS}/2+n)(m_{AdS}/2+n-1)}]^{1/2}c(n_1,n_2-1,n+1)
\end{aligned}\tag{9.14}$$

Consider first the case of massive and massless particles. As noted in Sec. 1.3, the contraction to the Poincare invariant case can be performed as follows. If R is a parameter with the dimension *length* and the operators P_μ ($\mu=0,1,2,3$) are defined as $P_\mu = M_{\mu 4}/2R$ then in the formal limit when $R\to\infty$, $M_{\mu 4}\to\infty$ but the ratio $M_{\mu 4}/R$ remains finite, one gets the commutation relations of the Poincare algebra from the commutation relations of the so(2,3) algebra. Therefore in situations where Poincare limit is valid with a high accuracy, the operators $M_{\mu 4}$ are much greater than the other operators. The quantum numbers (m_{AdS},n_1,n_2,n) should be very large since in the formal limit $R\to\infty$, $m_{AdS}/2R$ should become the standard Poincare mass and the quantities $(n_1/2R, n_2/2R, n/2R)$ should become continuous momentum variables.

A typical form of the semiclassical wave function is

$$c(n_1,n_2,n) = a(n_1,n_2,n)exp[i(n_1\varphi_1+n_2\varphi_2+n\varphi)]$$

where the amplitude $a(n_1,n_2,n)$ has a sharp maximum at semiclassical values of (n_1,n_2,n). Since the numbers (n_1,n_2,n) are very large, when some of them change by one, the major change of $c(n_1,n_2,n)$ comes from the rapidly oscillating exponent. As a consequence, in semiclassical approximation each representation operator becomes the operator of multiplication by a function and, as follows from Eqs. (8.15,9.14)

$$M_{04} = m_{AdS} + 2(n_1 + n_2 + n) \quad M_{12} = 2(n_1 - n_2)$$
$$M_{10} = 2[n_1(m_{AdS}/2 + n + n_1)]^{1/2} sin\varphi_1 - 2[n_2(m_{AdS}/2 + n + n_2)]^{1/2} sin\varphi_2$$
$$M_{20} = 2[n_1(m_{AdS}/2 + n + n_1)]^{1/2} cos\varphi_1 + 2[n_2(m_{AdS}/2 + n + n_2)]^{1/2} cos\varphi_2$$
$$M_{14} = -2[n_1(m_{AdS}/2 + n + n_1)]^{1/2} cos\varphi_1 + 2[n_2(m_{AdS}/2 + n + n_2)]^{1/2} cos\varphi_2$$
$$M_{24} = 2[n_1(m_{AdS}/2 + n + n_1)]^{1/2} sin\varphi_1 + 2[n_2(m_{AdS}/2 + n + n_2)]^{1/2} sin\varphi_2$$
$$M_{23} = 2\frac{[n(m_{AdS}+n)]^{1/2}}{m_{AdS}/2+n}\{[n_1(m_{AdS}/2 + n + n_2)]^{1/2} cos(\varphi - \varphi_1) +$$
$$[n_2(m_{AdS}/2 + n + n_1)]^{1/2} cos(\varphi - \varphi_2)\}$$
$$M_{31} = 2\frac{[n(m_{AdS}+n)]^{1/2}}{m_{AdS}/2+n}\{[n_1(m_{AdS}/2 + n + n_2)]^{1/2} sin(\varphi - \varphi_1) -$$
$$[n_2(m_{AdS}/2 + n + n_1)]^{1/2} sin(\varphi - \varphi_2)\}$$
$$M_{34} = 2\frac{[n(m_{AdS}+n)]^{1/2}}{m_{AdS}/2+n}\{[(m_{AdS}/2 + n + n_1)(m_{AdS}/2 + n + n_2)]^{1/2} cos\varphi +$$
$$(n_1 n_2)^{1/2} cos(\varphi - \varphi_1 - \varphi_2)\}$$
$$M_{30} = -2\frac{[n(m_{AdS}+n)]^{1/2}}{m_{AdS}/2+n}\{[(m_{AdS}/2 + n + n_1)(m_{AdS}/2 + n + n_2)]^{1/2} sin\varphi -$$
$$(n_1 n_2)^{1/2} sin(\varphi - \varphi_1 - \varphi_2)\} \quad (9.15)$$

We now consider what restrictions follow from the fact that in Poincare limit the operators $M_{\mu 4}$ $(\mu = 0, 1, 2, 3)$ should be much greater than the other operators. The first conclusion is that, as follows from the first expression in Eq. (9.15), the quantum numbers n_1 and n_2 should be such that $|n_1 - n_2| \ll n_1, n_2$. Therefore in the main approximation in $1/R$ we have that $n_1 \approx n_2$. Then it follows from the last expression that $sin\varphi$ should be of the order of $1/R$ and hence φ ahould be close either to zero or to π. Then it follows from the last four expressions in Eq. (9.15) that the operators $M_{\mu 4}$ will be indeed much greater than the other operators if $\varphi_2 \approx \pi - \varphi_1$ and in the main approximation in $1/R$

$$M_{04} = m_{AdS} + 2(2n_1 + n), \quad M_{14} = -4[n_1(m_{AdS}/2 + n + n_1)]^{1/2} cos\varphi_1$$
$$M_{14} = 4[n_1(m_{AdS}/2 + n + n_1)]^{1/2} sin\varphi_1, \quad M_{34} = \pm 2[n(m_{AdS} + n)]^{1/2} \quad (9.16)$$

where M_{34} is positive if φ is close to zero and negative if φ is close to π. In this approximation we have that $M_{04}^2 - \sum_{i=1}^{3} M_{i4}^2 = m_{AdS}^2$ which ensures that in Poincare limit we have the correct relation between the energy and momentum.

Consider now the case of the spinless Rac singleton. Then $m_{AdS} = 1$ and the quantity n can take only the values 0 and 1. Since the expressions in Eq. (9.14) are exact, we can use them in the given case as well. However, since the quantum number n cannot be large, we now cannot consider the n dependence of the wave function in semiclassical approximation. At the same time, if the numbers (n_1, n_2) are very large, the dependence of the wave function on (n_1, n_2) still can be considered in this approximation assuming that the wave function contains the rapidly oscillating exponent $exp[i(n_1\varphi_1+n_2\varphi_2)]$. Hence Eq. (9.15) remains valid but for calculating the operators M_{a3} $(a=0,1,2,4)$ one can use the fact that, as follows from Eq. (9.14)

$$\begin{aligned} b"c(n_1,n_2,n) &= 2(n_1n_2)^{1/2}\{c(n_1,n_2,0)\delta_{n1}+exp[-i(\varphi_1+\varphi_2)]c(n_1,n_2,1)\delta_{n0}\} \\ b'c(n_1,n_2,n) &= 2(n_1n_2)^{1/2}\{c(n_1,n_2,1)\delta_{n0}+exp[i(\varphi_1+\varphi_2)]c(n_1,n_2,0)\delta_{n1}\} \\ L_+c(n_1,n_2,n) &= 2(n_1n_2)^{1/2}\{exp(-i\varphi_1)c(n_1,n_2,1)\delta_{n0}+exp(i\varphi_2)c(n_1,n_2,0)\delta_{n1}\} \\ L_-c(n_1,n_2,n) &= 2(n_1n_2)^{1/2}\{exp(i\varphi_1)c(n_1,n_2,0)\delta_{n1}+ \\ &\quad exp(-i\varphi_2)c(n_1,n_2,1)\delta_{n0}\} \end{aligned} \tag{9.17}$$

where δ is the Kronecker symbol. Then the mean values of these operators can be written as

$$\begin{aligned} &< b" >= A\{exp(i\varphi)+exp[-i(\varphi+\varphi_1+\varphi_2)]\}, \quad < b' >=< b" >^* \\ &< L_+ >= A\{exp[-i(\varphi+\varphi_1)]+exp[i(\varphi+\varphi_2)]\}, \quad < L_- >=< L_+ >^* \end{aligned} \tag{9.18}$$

where

$$\sum_{n_1,n_2} 2(n_1n_2)^{1/2}c(n_1,n_2,1)^*c(n_1,n_2,0)=Aexp(i\varphi)$$

and we use * to denote the complex conjugation. By analogy with the above discussion, we conclude that the Poincare limit exists only if $\varphi_2\approx\pi-\varphi_1$ and φ is close either to zero or π. Then

$$M_{04}\approx 4n_1, \quad M_{14}\approx -4n_1cos(\varphi_1), \quad M_{24}\approx 4n_1sin(\varphi_1) \tag{9.19}$$

and the mean value of the operator M_{34} is much less than M_{14} and M_{24}.

Consider now the case of the Di singleton. It is characterized by $q_1=3/2$, $q_2=1/2$. Then, as shown in the preceding sections, $s=1$, the quantum number n can take only the value $n=0$ and the quantum number k can take only the values $k=0,1$. We denote $e_0=e(n=0,k=0)$ and $e_1=e(n=0,k=1)$. Then, as shown in the preceding section, $e_1=L_-e_0$ and the basis of the IR in standard theory consists of elements $e_0(n_1,n_2)=(a_1")^{n_1}(a_2")^{n_2}e_0$ and $e_1(n_1,n_2)=(a_1")^{n_1}(a_2")^{n_2}e_1$ $(n_1,n_2=0,1,...\infty)$.

As explained in Sec. 8.3, $e(n=1,k=0)$ should be defined as $[b"(h_1-1)-a_1"L_-]e_0$ and $e(n=1,k=1)$ should be defined as $[b"(h_2-1)-a_2"L_+]e_1$. Since in the case of the Di singleton $e(n=1,k=0)=e(n=1,k=1)=0$, it follows from Eq. (8.12) that

$$L_+e_0=L_-e_1=0, \quad L_-e_0=e_1 \quad L_+e_1=e_0, \quad b"e_0=a_1"e_1, \quad b"e_1=a_2"e_0 \tag{9.20}$$

Now it follows from Eq. (8.14) that

$$\begin{aligned} &b"e_0(n_1,n_2)=e_1(n_1+1,n_2), \quad b"e_1(n_1,n_2)=e_0(n_1,n_2+1) \\ &b'e_0(n_1,n_2)=(n_1+1)n_2e_1(n_1,n_2-1), \quad b'e_1(n_1,n_2)=n_1(n_2+1)e_0(n_1-1,n_2) \\ &L_+e_0(n_1,n_2)=n_2e_1(n_1+1,n_2-1), \quad L_+e_1(n_1,n_2)=(n_2+1)e_0(n_1,n_2) \\ &L_-e_0(n_1,n_2)=(n_1+1)e_1(n_1,n_2), \quad L_-e_1(n_1,n_2)=n_1e_0(n_1-1,n_2+1) \end{aligned} \tag{9.21}$$

As follows from Eqs. (8.8) and (8.12)

$$||e_0(n_1,n_2)|| = (n_1+1)!n_2!(n_2+1)^{1/2}, \quad ||e_1(n_1,n_2)|| = n_1!(n_2+1)!(n_1+1)^{1/2} \quad (9.22)$$

Hence one can define the normalized basis elements $\tilde{e}_j(n_1,n_2)$ $(j=0,1)$ and any element in the representation space can be written as $x=\sum_{j=0}^{1} c_j(n_1,n_2)\tilde{e}_j(n_1,n_2)$. By analogy with the above discussion, one can show that a necessary condition for the Poincare limit in semiclassical approximation is that the quantities (n_1,n_2) are very large, $n_1\approx n_2$, the functions $c_j(n_1,n_2)$ contain a rapidly oscillating exponents $exp[i(n_1\varphi_1+n_2\varphi_2)]$ and $\varphi_2\approx \pi-\varphi_1$. In this approximation one can obtain the results given by Eq. (9.15) while calculating the operators M_{a3} $(a=0,1,2,4)$ can be performed as follows.

One can represent the wave function as $(c_0(n_1,n_2),c_1(n_1,n_2))$ and then, as follows from Eqs. (9.21) and (9.22)

$$\begin{aligned} b"(c_0(n_1,n_2),c_1(n_1,n_2)) &\approx n_1(exp(-i\varphi_1)c_1(n_1,n_2),exp(-i\varphi_2)c_0(n_1,n_2))\\ b'(c_0(n_1,n_2),c_1(n_1,n_2)) &\approx n_1(exp(i\varphi_2)c_1(n_1,n_2),exp(i\varphi_1)c_0(n_1,n_2))\\ L_+(c_0(n_1,n_2),c_1(n_1,n_2)) &\approx n_1(exp[-i(\varphi_1-\varphi_2)]c_1(n_1,n_2),c_0(n_1,n_2))\\ L_-(c_0(n_1,n_2),c_1(n_1,n_2)) &\approx n_1(c_1(n_1,n_2),exp[i(\varphi_1-\varphi_2)]c_0(n_1,n_2)) \end{aligned} \quad (9.23)$$

Now it follows from Eqs. (8.15) and (9.23) that the mean values of the operators M_{a3} are given by

$$\begin{aligned} &< M_{34} >\approx 2A[cos(\varphi-\varphi_1)+cos(\varphi+\varphi_2)]\\ &< M_{30} >\approx 2A[sin(\varphi-\varphi_1)-sin(\varphi+\varphi_2)]\\ &< M_{23} >\approx 2A[cos(\varphi-\varphi_1+\varphi_2)+cos\varphi]\\ &< M_{31} >\approx 2A[sin(\varphi-\varphi_1+\varphi_2)-sin\varphi] \end{aligned} \quad (9.24)$$

where

$$\sum_{n_1n_2} n_1c_1(n_1,n_2)^*c_0(n_1,n_2) = Aexp(i\varphi)$$

If $\varphi_2\approx\pi-\varphi_1$ then it is easy to see that the Poincare limit for $< M_{23} >$ and $< M_{31} >$ exists if $\varphi\approx\varphi_1$ or $\varphi\approx\varphi_1+\pi$. In that case the Poincare limit for $< M_{34} >$ and $< M_{30} >$ exists as well and $< M_{34} >$ disappears in the main approximation.

We have shown that if the operators M_{ab} are defined by Eq. (8.15) then in Poincare limit the z component of the momentum is negligible for both, the Di and Rac singletons. This result could be expected from Eq. (9.16) since for them neither m_{AdS} nor n can be large numbers. As noted in the remark after Eq. (8.15), the definition (8.15) is not unique and, in particular, any definition obtained from Eq. (8.15) by cyclic permutation of the indices $(1,2,3)$ is valid as well. Therefore we conclude that in standard theory, the Di and Rac singletons have the property that in the Poincare limit they are characterized by two independent components of the momentum, not three as usual particles. This is a consequence of the fact that for singletons only the quantum numbers n_1 and n_2 can be very large.

The properties of singletons in Poincare limit have been discussed by several authors, and their conclusions are not in agreement with each other (a detailed list of references can be found e.g. in Refs. [115, 116]). In particular, there are statements that the Poincare limit for singletons does not exist or that in this limit all the components of the four-momentum become zero. The above consideration shows that Poincare limit for singletons can be investigated in full analogy with Poincare limit for usual particles. In particular, the statement that the singleton energy in Poincare limit becomes zero is not in agreement with the fact that each massless particle (for which the energy in Poincare limit is not zero) can be represented as a composite state of two singletons. The fact that the standard singleton momentum can have only two independent components does not contradict the fact that the momentum of a massless particle has three independent components since, as noted above, the independent momentum components of two singletons can be in different planes.

9.4 Tensor products of singleton IRs

We now return to the presentation when the properties of singletons in standard and modular approaches are discussed in parallel. The tensor products of singleton IRs have been defined in Sec. 9.1. If $e^{(j)}(n_1^{(j)}, n_2^{(j)}, n^{(j)}, k^{(j)})$ $(j = 1, 2)$ are the basis elements of the IR for singleton j then the basis elements in the representation space of the tensor product can be chosen as

$$e(n_1^{(1)}, n_2^{(1)}, n^{(1)}, k^{(1)}, n_1^{(2)}, n_2^{(2)}, n^{(2)}, k^{(2)}) = e^{(1)}(n_1^{(1)}, n_2^{(1)}, n^{(1)}, k^{(1)}) \times e^{(2)}(n_1^{(2)}, n_2^{(2)}, n^{(2)}, k^{(2)}) \quad (9.25)$$

In the case of the tensor product of singleton IRs of different types, we assume that singleton 1 is Di and singleton 2 is Rac.

Consider a vector

$$e(q) = \sum_{i=0}^{q} c(i, q) e^{(1)}(i, 0, 0, 0) \times e^{(2)}(q - i, 0, 0, 0) \quad (9.26)$$

where the coefficients $c(i, q)$ are given by Eq. (9.5) such that the $q_0^{(j)}$ should be replaced by $q_1^{(j)}$ $(j = 1, 2)$. Since $h_2^{(j)} e^{(j)}(i, 0, 0, 0) = ((p+1)/2) e^{(j)}(i, 0, 0, 0)$ $(j = 1, 2)$ then the vector $e(q)$ is the eigenvector of the operator $h_2 = h_2^{(1)} + h_2^{(2)}$ with the eigenvalue $q_2 = 1$ and satisfies the condition $a_2' e(q) = 0$ where $a_2' = a_2^{(1)'} + a_2^{(2)'}$. As follows from the results of Sec. 9.2, $e(q)$ is the eigenvector of the operator $h_1 = h_1^{(1)} + h_1^{(2)}$ with the eigenvalue $q_1 = q_1^{(1)} + q_1^{(2)} + 2q$ and satisfies the condition $a_1' e(q) = 0$ where $a_1' = a_1^{(1)'} + a_1^{(2)'}$. It is obvious that the value of q_1 equals $3 + 2q$ for the tensor product $Di \times Di$, $2+2q$ for the tensor product $Di \times Rac$ and $1+2q$ for the tensor product $Rac \times Rac$.

As follows from Eqs. (8.14) and (8.27), in the case of IRs

$$\begin{aligned}
&b'e(n_1n_2nk) = [(a_1")^{n_1}(a_2")^{n_2}b' + n_1(a_1")^{n_1-1}(a_2")^{n_2}L_+ + \\
&n_2(a_1")^{n_1}(a_2")^{n_2-1}L_- + n_1n_2(a_1")^{n_1-1}(a_2")^{n_2-1}b"]e(0,0,n,k) \\
&L_+e(n_1n_2nk) = [(a_1")^{n_1}(a_2")^{n_2}L_+ + n_2(a_1")^{n_1}(a_2")^{n_2-1}b"]e(0,0,n,k)
\end{aligned} \quad (9.27)$$

Therefore, $e(q)$ satisfies the conditions $b'e(q) = L_+e(q) = 0$ where $b' = b^{(1)'} + b^{(2)'}$ and $L_+ = L_+^{(1)} + L_+^{(2)}$. Hence, $e(q)$ is an analog of the vector e_0 in Eq. (8.19) and generates an IR corresponding to the quantum numbers $(q_1, q_2 = 1)$.

We conclude that the tensor product of singleton IRs contains massless IRs corresponding to $q_1 = q_1^{(1)} + q_1^{(2)} + 2q$. As follows from the results of Sect. 9.2 (see the remark after Eq. (9.5)), q can take the values $0, 1, ..., p - q_1^{(1)}$. Therefore $Rac \times Rac$ contains massless IRs with $s = 0, 2, 4, ..., (p-1)$, $Di \times Rac$ contains massless IRs with $s = 1, 3, 5, ...(p-2)$ and $Di \times Di$ contains massless IRs with $s = 2, 4, ...(p-1)$. In addition, as noted in Ref. [113], $Di \times Di$ contains a spinless massive IR corresponding to $q_1 = q_2 = 2$. This question will be discussed in Sec. 9.6

Our next goal is to investigate whether or not all those IRs give a complete decomposition of the corresponding tensor products. For example, as follows from Eq. (8.35), for the product $Rac \times Rac$ this would be the case if the sum $\sum_{k=0}^{(p-1)/2} Dim(2k)$ equals $(p^2+1)^2/4 = p^4/4 + O(p^2)$. However, as follows from Eqs. (8.37) and (8.39), this sum can be easily estimated as $11p^4/48 + O(p^3)$ and hence, in contrast to the Flato-Fronsdal result in standard theory, in the modular case the decomposition of $Rac \times Rac$ contains not only massles IRs. Analogously, the sum of dimensions of massless IRs entering into the decompositions of $Di \times Rac$ and $Di \times Di$ also can be easily estimated as $11p^4/48 + O(p^3)$ what is less than $p^4/4 + O(p^2)$. The reason is that in the modular case the decompositions of the tensor products of singletons contain not only massles IRs but also special IRs. We will not investigate the modular analog of the Flato-Fronsdal theorem [113] but concentrate our efforts on finding a full solution of the problem in the supersymmetric case.

9.5 Supersingleton IR

In this section we consider the supersingleton IR exclusively in terms of the fermionic operators without decomposing the IR into the Di and Rac IRs. As a preparatory step, we first consider IRs of a simple superalgebra generated by two fermionic operators $(d', d")$ and one bosonic operator h such that

$$h = \{d', d"\}, \quad [h, d'] = -d', \quad [h, d"] = d" \quad (9.28)$$

Here the first expression shows that, by analogy with the osp(1,4) superalgebra, the relations (9.28) can be formulated only in terms of the fermionic operators.

Consider an IR of the algebra (9.28) generated by a vector e_0 such that

$$d'e_0 = 0, \quad d'd"e_0 = q_0 e_0 \tag{9.29}$$

and define $e_n = (d")^n e_0$. Then $d'e_n = a(n)e_{n-1}$ where, as follows from Eq. (9.29), $a(0) = 0$, $a(1) = q_0$ and $a(n) = q_0 + n - 1 - a(n-1)$. It is easy to prove by induction that

$$a(n) = \frac{1}{2}\{(q_0 - \frac{1}{2})[1 - (-1)^n] + n\} \tag{9.30}$$

The maximum possible value of n can be found from the condition that $a(n_{max}) \neq 0$, $a(n_{max}+1) = 0$. In the special case of the supersingleton, we will be interested in the case when $q_0 = (p+1)2$. Then, as follows from Eq. (9.30), $a(n) = n/2$. Therefore $n_{max} = p - 1$ and the dimension of the IR is p. In the general case, if $q_0 \neq 0$ then $a(n) = 0$ if $n = 2p + 1 - 2q_0$ and the dimension of the IR is $D(q_0) = 2p + 1 - 2q_0$.

Consider now the supersingleton IR. Let $x = (d_1^" d_2^" - d_2^" d_1^")e_0$. Then, as follows from Eq. (8.93), $d_1'x = (2q_1 - 1)d_2^" e_0$ and $d_2'x = (1 - 2q_2)d_1^" e_0$. Since $q_1 = q_2 = (p+1)/2$ we have that $d_1'x = d_2'x = 0$ and therefore $x = 0$. Hence the actions of the operators $d_1^"$ and $d_2^"$ on e_0 commute with each other. If n is even then $d_1^"(d_2^")^n e_0 = (d_2^")^n d_1^" e_0$ as a consequence of Eq. (8.93) and if n is odd then $d_1^"(d_2^")^n e_0 = (d_2^")^{n-1} d_1^" d_2^" e_0 = (d_2^")^n d_1^" e_0$ in view of the fact that $x = 0$. Analogously one can prove that $d_2^"(d_1^")^n e_0 = (d_1^")^n d_2^" e_0$. We now can prove that $d_1^"(d_2^")^n (d_1^")^k e_0 = (d_2^")^n (d_1^")^{k+1} e_0$. Indeed, if n is even, this is obvious while if n is odd then

$$d_1^"(d_2^")^n(d_1^")^k e_0 = (d_2^")^{n-1} d_1^" d_2^" (d_1^")^k e_0 = (d_2^")^{n-1}(d_1^")^{k+1} d_2^" e_0 = (d_2^")^n (d_1^")^{k+1} e_0$$

and analogously $d_2^"(d_1^")^n(d_2^")^k e_0 = (d_1^")^n(d_2^")^{k+1} e_0$. Therefore the supersingleton IR is distinguished among other IRs of the osp(1,4) superalgebra by the fact that the operators $d_1^"$ and $d_2^"$ commute in the representation space of this IR. Hence the basis of the representation space can be chosen in the form $e(nk) = (d_1^")^n (d_2^")^k e_0$. As a consequence of the above consideration, $n, k = 0, 1, ...p - 1$ and the dimension of the IR is p^2 in agreement with Eq. (8.35).

The above results can be immediately generalized to the case of higher dimensions. Consider a superalgebra defined by the set of operators $(d_j', d_j")$ where $j = 1, 2, ...J$ and, by analogy with Eq. (8.93), any triplet of the operators (A, B, C) satisfies the commutation-anticommutation relation

$$[A, \{B, C\}] = F(A, B)C + F(A, C)B \tag{9.31}$$

where the form $F(A, B)$ is skew symmetric, $F(d_j', d_j") = 1$ $(j = 1, 2, ...J)$ and the other independent values of $F(A, B)$ are equal to zero. The higher-dimensional analog of the supersingleton IR can now be defined such that the representation space contains a vector e_0 satisfying the conditions

$$d_j'e_0 = 0, \quad d_j'd_j"e_0 = 1/2 \quad (j = 1, 2, ...J) \tag{9.32}$$

The basis of the representation space can be chosen in the form $e(n_1, n_2, ...n_J) = (d_1")^{n_1}(d_2")^{n_2}\cdots(d_J")^{n_J}e_0$. In full analogy with the above consideration one can show that the operators $(d_1", ...d_J")$ mutually commute on the representation space. As a consequence, in the modular case each of the numbers n_j $(j = 1, 2, ...J)$ can take the values $0, 1, ...p-1$ and the dimension of the IR is p^J. The fact that singleton physics can be directly generalized to the case of higher dimensions has been indicated by several authors (see e.g. Ref. [115] and references therein).

9.6 Tensor product of supersingleton IRs

We first consider the tensor product of IRs of the superalgebra (9.28) with $q_0 = (p+1)/2$. The representation space of the tensor product consists of all linear combinations of elements $x^{(1)} \times x^{(2)}$ where $x^{(j)}$ is an element of the representation space for the IR j $(j = 1, 2)$. The representation operators of the tensor product are linear combinations of the operators $(d', d")$ where $d' = d^{(1)'} + d^{(2)'}$ and $d" = d^{(1)"} + d^{(2)"}$. Here $d^{(j)'}$ and $d^{(j)"}$ mean the operators acting in the representation spaces of IRs 1 and 2, respectively. In contrast to the case of tensor products of IRs of the sp(2) and so(2,3) algebras, we now require that if $d^{(j)}$ is some of the d-operators for the IR j then the operators $d^{(1)}$ and $d^{(2)}$ anticommute rather than commute, i.e. $\{d^{(1)}, d^{(2)}\} = 0$ Then it is obvious that the independent operators defining the tensor product satisfy Eq. (9.28).

Let $e_0^{(j)}$ be the generating vector for IR j and $e_i^{(j)} = (d^{(j)"})^i e_0^{(j)}$. Consider the following element of the representation space of the tensor product

$$e(k) = \sum_{i=0}^{k} c(i)(e_i^{(1)} \times e_{k-i}^{(2)}) \tag{9.33}$$

where $c(i)$ is some function. This element will be the generating vector of the IR of the superalgebra (9.28) if $d'e(q) = 0$. As follows from the above results and Eq. (9.33)

$$d'e(q) = \frac{1}{2}\sum_{i=1}^{k} ic(i)(e_{i-1}^{(1)} \times e_{k-i}^{(2)}) + \frac{1}{2}\sum_{i=0}^{k-1}(-1)^i(k-i)c(i)(e_i^{(1)} \times e_{k-i-1}^{(2)}) \tag{9.34}$$

Therefore $d'e(k) = 0$ is satisfied if $k = 0$ or

$$c(i+1) = (-1)^{i+1}\frac{k-i}{i+1}c(i), \quad i = 0, 1, ...k-1 \tag{9.35}$$

when $k \neq 0$. As follows from this expression, if $c(0) = 1$ then

$$c(i) = (-1)^{\frac{i(i+1)}{2}} C_k^i \tag{9.36}$$

where $C_k^i = k!/i!(k-i)!$ is the binomial coefficient. As follows from Eq. (9.30), the possible values of k are $0, 1, ...p-1$ and, as follows from Eq. (9.33), $he(k) = q_0 e(k)$ where $q_0 = 1 + k$. The fact that the tensor product is fully decomposable into IRs with the different values of k follows from the relation $\sum_{q_0=1}^{p} D(q_0) = p^2$.

The tensor product of the supersingleton IRs can be constructed as follows. The representation space of the tensor product consists of all linear combinations of elements $x^{(1)} \times x^{(2)}$ where $x^{(j)}$ is an element of the representation space for the supersingleton j $(j = 1, 2)$. The fermionic operators of the representation are linear combinations of the operators $(d_1', d_2', d_1", d_2")$ where $d_1' = d_1^{(1)'} + d_1^{(2)'}$ and analogously for the other operators. Here $d_k^{(j)'}$ and $d_k^{(j)"}$ $(k = 1, 2)$ mean the operators d_k' and $d_k"$ acting in the representation spaces of supersingletons 1 and 2, respectively. We also assume that if $d^{(j)}$ is some of the d-operators for supersingleton j then $\{d^{(1)}, d^{(2)}\} = 0$. Then all the d-operators of the tensor product satisfy Eq. (8.93) and the action of the bosonic operators in the tensor product can be defined by Eq. (8.94).

Let $e_0^{(j)}$ be the generating vector for supersingleton j (see Eq. (8.96)) and $e_0 = e_0^{(1)} \times e_0^{(2)}$. Consider the following element of the representation space of the tensor product:

$$x(k_1, k_2) = \sum_{i=0}^{k_1} \sum_{j=0}^{k_2} (-1)^{[\frac{i(i+1)}{2} + \frac{j(j+1)}{2} + k_1 j]} C_{k_1}^i C_{k_2}^j$$
$$(d_1^{(1)"})^i (d_1^{(2)"})^{k_1 - i} (d_2^{(1)"})^j (d_2^{(1)"})^{k_2 - j} e_0 \quad (k_1, k_2 = 0, 1, ...p-1) \qquad (9.37)$$

By using Eq. (8.93) and the results of this section, one can explicitly verify that all the $x(k_1, k_2)$ are the nonzero vectors and

$$d_1' x(k_1, k_2) = d_2' x(k_1, k_2) = 0, \quad d_2' d_1" x(k_1, k_2) = x(k_1 + 1, k_2 - 1) \qquad (9.38)$$

Since the $e_0^{(j)}$ $(j = 1, 2)$ are the generating vectors of the IRs of the osp(1,4) superalgebra with $(q_1, q_2) = ((p+1)/2, (p+1)/2)$, it follows from Eq. (8.95) that $x(k_1, k_2)$ is the generating vector of the IRs of the osp(1,4) superalgebra with $(q_1, q_2) = (1 + k_1, 1 + k_2)$ if $d_2' d_1" x(k_1, k_2) = 0$. Therefore, as follows from Eq. (9.38), this is the case if $k_2 = 0$. Hence the tensor product of the supersingleton IRs contains IRs of the osp(1,4) algebra corresponding to $(q_1, q_2) = (1 + k_1, 1)$ $(k_1 = 0, 1, ...p-1)$. As noted in Sect. 8.10, the case $(0, 1)$ can be treated either as the massless IR with $s = p - 1$ or as the special massive IR; the case $(1, 1)$ can be treated as the massive IR of the osp(1,4) superalgebra and the cases when $k_1 = 1, ...p-2$ can be treated as massless IRs with $s = k_1$.

The results of standard theory follow from the above results in the formal limit $p \to \infty$. Therefore in standard theory the decomposition of tensor product of supersingletons contains the IRs of the osp(1,4) superalgebra corresponding to $(q_1, q_2) = (1, 1), (2, 1), ...(\infty, 1)$ in agreement with the results obtained by Flato and Fronsdal [113] and Heidenreich [114].

As noted in Sect. 9.4, the Flato-Fronsdal result for the tensor product $Di \times Di$ is that it also contains a massive IR corresponding to $q_1 = q_2 = 2$. In terms of the fermionic operators this result can be obtained as follows. If $y = (d_1^{(1)"} d_2^{(2)"} - d_2^{(1)"} d_1^{(2)"})e_0$ then, as follows from Eqs. (8.93) and (8.94),

$$d_1^{(1)'} y = \frac{p+1}{2} d_2^{(2)"} e_0, \quad d_1^{(2)'} y = \frac{p+1}{2} d_2^{(1)"} e_0, \quad d_2^{(1)'} y = -\frac{p+1}{2} d_1^{(2)"} e_0$$
$$d_2^{(2)'} y = -\frac{p+1}{2} d_1^{(1)"} e_0, \quad h_1 y = h_2 y = 2y, \quad L_+ y = L_- y = 0 \tag{9.39}$$

Since $a_j' = (d_j')^2$ for $j = 1, 2$ (see Eq. (8.94)), it follows from these expressions that $a_1' y = a_2' y = 0$, i.e. y indeed is the generating vector for the IR of the so(2,3) algebra characterized by $q_1 = q_2 = 2$. However, y is not a generating vector for any IR of the osp(1,4) superalgebra since it does not satisfy the condition $d_1' y = d_2' y = 0$.

The vector $x(k_1, k_2)$ defined by Eq. (9.37) becomes the null vector when $k_1 = p$. Indeed, since $C_{k_1}^i = k_1!/[i!(k_1 - i)!]$, the sum over i in Eq. (9.37) does not contain terms with $i \neq 0$ and $i \neq p$. At the same time, if $i = 0$ or $i = p$ the corresponding terms are also the null vectors since, as follows from the results of the preceding section, $(d_1')^p e_0 = (d_2')^p e_0 = 0$. It is obvious that this result is valid only in the modular case and does not have an analog in standard theory. Therefore, as follows from Eq. (9.38), the decomposition of the tensor products of two supersingletons also contains IRs of the osp(1,4) superalgebra characterized by $(q_1, q_2) = (0, 0), (0, 1), (0, 2), ...(0, p-1)$.

We have shown that the decomposition of the tensor products of two supersingletons contains IRs of the osp(1,4) superalgebra characterized by the following values of (q_1, q_2):

$$(0, 0), (0, 1), (0, 2), ...(0, p-1), (1, 1), (2, 1), ...(p-1, 1)$$

The question arises whether this set of IRs is complete, i.e. the decomposition of the tensor products of two supersingletons does not contain other IRs of the osp(1,4) superalgebra. Since the dimension of the supersigleton IR is p^2 (see the preceding section), this is the case if

$$\sum_{k=0}^{p-1} SDim(0, k) + \sum_{k=1}^{p-1} SDim(1, k) = p^4 \tag{9.40}$$

It is obvious that $SDim(0, 0) = 1$ since the IR characterized by $(q_1, q_2) = (0, 0)$ is such that all the representation operators acting on the generating vector give zero. Therefore, as follows from Eq. (8.103), the condition (9.40) can be rewritten as

$$2 + Dim(0) + Dim(2, 2) + 2\sum_{s=1}^{p-2} Dim(s) + 2\sum_{q_2=1}^{p-1} Dim(0, q_2) = p^4 \tag{9.41}$$

since $Dim(1,1) = Dim(0)$. The expressions for $Dim(s)$ and $Dim(0,q_2)$ are given in Eqs. (8.37-8.40) and hence the only quantity which remains to be calculated is $Dim(2,2)$.

The IR of the so(2,3) algebra characterized by $(q_1,q_2) = (2,2)$ is the massive IR with $m_{AdS} = 4$ and $s = 0$. Therefore, as follows from the results of Sect. 8.2, the quantity k in Eq. (4.4) can take only the value $k = 0$ and the quantity n can take the values $0,1,...n_{max}$ where $n_{max} = p-2$. Hence, as follows from Eqs. (8.7) and (8.28)

$$Dim(2,2) = \sum_{n=0}^{p-2}(p-1-n)^2 = \frac{1}{6}p(p-1)(2p-1) \qquad (9.42)$$

The validity of Eq. (9.41) now follows from Eqs. (8.37-8.40,9.42).

The main result of this chapter can now be formulated as follows:

In a quantum theory over a Galois field, the tensor product of two Dirac supersingletons is fully decomposable into the following IRs of the osp(1,4) superalgebra:

- *Massive IR characterized by* $(q_1 = 1, q_2 = 1)$
- *Massless IRs characterized by* $(q_1 = 2,...p-1, q_2 = 1)$
- *Special IRs characterized by* $(q_1 = 0, q_2 = 0,1,...p-1)$

and the multiplicity of each IR in the decomposition is equal to one.

Chapter 10

Discussion and conclusion

In Secs. 1.1 and 6.1 we argue that the main reason of crisis in physics is that nature, which is fundamentally discrete, is described by continuous mathematics. Moreover, no ultimate physical theory can be based on continuous mathematics because it is not self-consistent (as a consequence of Gödel's incompleteness theorems). In the first part of the work we discuss inconsistencies in standard approach to quantum theory and then we reformulate the theory such that it can be naturally generalized to a formulation based on discrete mathematics. In this chapter we discuss the main results of the present work in position operator, cosmological constant problem, gravity and particle theory.

10.1 Position operator and wave packet spreading

In standard physics education the position operator is typically discussed only in non-relativistic quantum mechanics. Here it is postulated that coordinate and momentum representations are related to each other by the Fourier transform and this leads to famous uncertainty relations. This postulate has been accepted from the beginning of quantum theory by analogy with classical electrodynamics. We argue that the postulate is based neither on strong theoretical arguments nor on experimental data.

In relativistic quantum theory local fields are discussed but typically in standard textbooks the argument x of those fields is not associated with any position operator (in spite of the principle of quantum theory that any physical quantity can be discussed only in conjunction with the operator of this quantity). Probably one of the reasons is that local quantum fields do not have a probabilistic interpretation and play only an auxiliary role for constructing the S-matrix in momentum space. When this construction is accomplished the theory does not contain space-time anymore in the spirit of the Heisenberg S-matrix program that in quantum theory one can descibe only transitions of states from the infinite past when $t \to -\infty$ to the distant future when $t \to +\infty$. As a consequence, many physicists believe that the position operators is meaningful only in nonrelativistic theory while in relativistic theory no position operator is needed.

However, relativistic position operator is needed in many problems. For example, when we consider how photons created on distant objects move to Earth we should know where those photons have been created (on Sun, Sirius or other objects), what is the (approximate) trajectory of those photons etc. Meanwhile many quantum physicists are not aware of the fact that relativistic position operator has been intensively discussed in papers by Newton and Wigner, Pryce, Hawton and other authors. By analogy with nonrelativistic quantum mechanics, in those papers the position and momentum operators are also related to each other by the Fourier transform.

Immediately after creation of quantum theory it has been realized that an inevitable consequence of the fact that the position and momentum operators are related to each other by the Fourier transform is the effect of wave packing spreading (WPS). Several well-known physicists (e.g. de Broglie) treated this fact as unacceptable and proposed alternative approaches to quantum theory. At the same time, it has not been shown that numerical results on WPS contradict experimental data. For example, it is known that for macroscopic bodies the effect of WPS is negligible and it is probably believed that in experiments on the Earth with atoms and elementary particles spreading does not have enough time to manifest itself.

However, it seems rather strange that no one has posed a problem of what happens to photons from distant stars which can travel to Earth even for billions of years. As shown in Chap. 2, the results for WPS calculated in standard theory are such that this effect is very important even for close stars and planets. As a consequence, in standard theory we have several striking paradoxes discussed in Chap. 2.

We propose a consistent construction of the position operator where the position and momentum operators are not related to each other by the Fourier transform. Then the effect of WPS in directions perpendicular to the particle momentum is absent and the paradoxes are resolved. Different components of the new position operator do not commute with each other and, as a consequence, there is no wave function in coordinate representation.

Our results give a strong arguments that the notion of space-time is pure classical and does not exist on quantum level. Hence fundamental quantum theory should not be based on Lagrangians and quantum field in coordinate representation.

10.2 Cosmological constant problem

As noted in Sect. 1.5, one of the main ideas of this work is that gravity might be not an interaction but simply a manifestation of de Sitter symmetry over a Galois field. This is obviously not in the spirit of mainstream approaches that gravity is a manifestation of the graviton exchange or holographic principle. Our approach does not involve General Relativity, quantum field theory (QFT), string theory, loop quantum gravity or other sophisticated theories. We consider only systems of *free* bodies in de Sitter invariant quantum mechanics.

Then the fact that we observe the cosmological repulsion is a strong argument that the de Sitter (dS) symmetry is a more pertinent symmetry than Poincare or anti de Sitter (AdS) ones. As shown in Refs. [37, 17] and in the present work, the phenomenon of the cosmological repulsion can be easily understood by considering semiclassical approximation in standard dS invariant quantum mechanics of two free bodies. In the framework of this consideration it becomes immediately clear that the cosmological constant problem does not exist and there is no need to involve empty space-time background, dark energy or other artificial notions. This phenomenon can be easily explained by using only standard quantum-mechanical notions without involving dS space, metric, connections or other notions of Riemannian geometry.

One might wonder why such a simple explanation has not been widely discussed in the literature. According to our observations, this is a manifestation of the fact that even physicists working on dS QFT are not familiar with basic facts about irreducible representations (IRs) of the dS algebra. It is difficult to imagine how standard Poincare invariant quantum theory can be constructed without involving well-known results on IRs of the Poincare algebra. Therefore it is reasonable to think that when Poincare invariance is replaced by dS one, IRs of the Poincare algebra should be replaced by IRs of the dS algebra. However, physicists working on QFT in curved space-time believe that fields are more fundamental than particles and therefore there is no need to involve IRs.

10.3 Gravity

The mainstream approach to gravity is such that gravity is the fourth (and probably the last) interaction which should be unified with electromagnetic, weak and strong interactions. While the electromagnetic interaction is a manifestation of the photon exchange, the weak interaction is a manifestation of the W and Z boson exchange and the strong interaction is a manifestation of the gluon exchange, gravity is supposed to be a manifestation of the graviton exchange. However, the notion of the exchange by virtual particles is taken from particle theory while gravity is known only at macroscopic level. Hence thinking that gravity can be explained by mechanisms analogous to those in particle theory is a great extrapolation.

There are several theoretical arguments in favor of the graviton exchange. In particular, in the nonrelativistic approximation Feynman diagrams for the graviton exchange can recover the Newton gravitational law by analogy with how Feynman diagrams for the photon exchange can recover the Coulomb law. However, the Newton gravitational law is known only on macroscopic level and, as noted in Sec. 2.1, the conclusion that the photon exchange reproduces the Coulomb law can be made only if one assumes that coordinate and momentum representations are related to each other by the Fourier transform. As discussed in Chaps. 1 and 2, on quantum level the coordinates are not needed and, as shown in Chap. 2, standard position operator contradicts experiments. In addition, as noted in Sec. 2.1, even on classical level the Coulomb law for pointlike electric charges has not been verified with a high accuracy. So on macroscopic level the validity of the Newton gravitation law has been verified with a much greater confidence than the Coulomb law. In view of these remarks, the argument that in quantum theory the Newton gravitational law should be obtained by analogy with the Coulomb law is not convincing.

The existence of gravitons can also be expected from the fact that GR (which is a classical theory) predicts the existence of gravitational waves and that from the point of view of quantum theory each classical wave should consist of particles. However, in spite of the fact that powerful facilities have been built for detecting gravitational waves, no unambiguous detections have been reported yet. In addition, as discussed in Sec. 5.8, the statement that the data on binary pulsars can be treated as an indirect confirmation of the existence of gravitational waves is strongly model dependent.

It has been also noted in Sec. 5.8 that any quantum theory of gravity can be tested only on macroscopic level. Hence, the problem is not only to construct quantum theory of gravity but also to understand a correct structure of the position operator on macroscopic level. However, in the literature the latter problem is not discussed because it is tacitly assumed that the position operator on macroscopic level is the same as in standard quantum theory. This is an additional great extrapolation which should be substantiated.

On the other hand, efforts to construct quantum theory of gravity have not been successful yet. Mainstream theories are based on the assumption that G is a fundamental constant while, as argued throughout this work, there are no solid reasons to think so. The assumption that G is a fundamental constant has been also adopted in GR. However, as discussed in Sec. 5.8, the existing results on non-Newtonian gravitational experiments cannot be treated as an unambiguous confirmation of GR.

In recent years a number of works has appeared where the authors treat gravity not as a fundamental interaction but as an emergent phenomenon. We believe that until the nature of gravity has been unambiguously understood, different approaches to gravity should be investigated. In the present work we consider gravity as a pure kinematical manifestation of quantum dS symmetry in semiclassical approximation.

In contrast to IRs of the Poincare and AdS algebras, in IRs of the dS algebra the particle mass *is not* the lowest eigenvalue of the dS Hamiltonian which has the spectrum in the range $(-\infty, \infty)$. As a consequence, the free mass operator of the two-particle system is not bounded below by $(m_1 + m_2)$ where m_1 and m_2 are the particle masses. The discussion in Secs. 3.6 and 5.1 shows that this property by no means implies that the theory is unphysical.

Since in Poincare and AdS invariant theories the spectrum of the free mass operator is bounded below by $(m_1 + m_2)$, in these theories it is impossible to obtain the correction $-Gm_1m_2/r$ to the mean value of this operator. However, in dS theory there is no law prohibiting such a correction. It is not a problem to indicate internal two-body wave functions for which the mean value of the mass operator contains $-Gm_1m_2/r$ with possible post-Newtonian corrections. The problem is to show that such wave functions are semiclassical with a high accuracy. As shown in Chaps. 3 and 5, in semiclassical approximation any correction to the standard mean value of the mass operator is negative and proportional to the energies of the particles. In particular, in the nonrelativistic approximation it is proportional to m_1m_2.

Our consideration in Chap. 5 gives additional arguments (to those posed in Chap. 2) that standard distance operator should be modified since a problem arises whether it is physical at macroscopic distances. In Chap. 5 we argue that it is not and propose a modification of the distance operator which has correct properties and gives for mean values of the free two-body mass operators the results compatible with Newton's gravity if the width of the de Sitter momentum distribution for a macroscopic body is inversely proportional to its mass. It has been also shown in Sec. 5.7 that for all known gravitational experiments, classical equations of motion can be obtained without involving the Lagrangian or Hamiltonian formalism but assuming only that time is defined as in Eq. (1.2), i.e. that the relation between the spatial displacement and the momentum is as in standard theory for free particles.

10.4 Quantum theory over a Galois field

In Chaps. 6 and 7 we argue that quantum theory should be based on Galois fields rather than complex numbers. We tried to make the presentation as simple as possible without assuming that the reader is familiar with Galois fields. Our version of a quantum theory over a Galois field (GFQT) gives a natural qualitative explanation why the width of the total dS momentum distribution of the macroscopic body is inversely proportional to its mass. In this approach neither G nor Λ can be fundamental physical constants. We argue that only $G\Lambda$ might have physical meaning. The calculation of this quantity is a very difficult problem since it requires a detailed knowledge of wave functions of many-body systems. However, GFQT gives clear indications that $G\Lambda$ contains a factor $1/lnp$ where p is the characteristic of the Galois field. We treat standard theory as a special case of GFQT in the formal limit $p \to \infty$. Therefore gravity disappears in this limit. Hence in our approach gravity is a consequence of the fact that dS symmetry is considered over a Galois field rather than the field of complex numbers.

In our approach gravity is a phenomenon which has a physical meaning only in situations when at least one body is macroscopic and can be described in the framework of semiclassical approximation. The result (5.29) shows that gravity depends on the width of the total dS momentum distributions for the bodies under consideration. However, when one mass is much greater than the other, the momentum distribution for the body with the lesser mass is not important. In particular, this is the case when one body is macroscopic and the other is the photon. At the same time, the phenomenon of gravity in systems consisting only of elementary particles has no physical meaning since gravity is not an interaction but simply a kinematical manifestation of dS invariance over a Galois field in semiclassical approximation. In this connection a problem arises what is the minimum mass when a body can be treated as macroscopic. This problem requires understanding of the structure of the many-body wave function.

Implications of GFQT in particle theory are discussed in the next sections.

10.5 Particle theory

10.5.1 Particle theory based on standard dS symmetry

As noted above, in standard theory (based on complex numbers) the fact that $\Lambda > 0$ is a strong indication that dS symmetry is more pertinent than Poincare and AdS symmetries. Hence it is reasonable to consider what happens when particle theory is considered from the point of view of dS symmetry. Then the key difference between IRs of the dS algebra on one hand and IRs of the Poincare and AdS algebras on the other is that in the former case one IR can be implemented only on the upper and lower Lorenz hyperboloids simultaneously. As a consequence, the number of states in IRs is always twice as big as the number of states in the corresponding IRs of the AdS or Poincare algebra. As explained in Sec. 3.5, an immediate consequence of this fact is that there are no neutral elementary particles in the theory.

Suppose that, by analogy with standard theory, one wishes to interpret states with the carrier on the upper hyperboloid as particles and states with the carrier on the lower hyperboloid as corresponding antiparticles. Then the first problem which arises is that the constant C in Eq. (3.58) is infinite and one cannot eliminate this constant by analogy with the AdS or Poincare theories. Suppose, however, that this constant can be eliminated at least in Poincare approximation where experiments show that the interpretation in terms of particles and antiparticles is physical. Then, as shown in Sec. 3.5, only fermions can be elementary.

One might think that theories where only fermions can be elementary and the photon (and also the graviton and the Higgs boson, if they exist) is not elementary, cannot be physical. However, several authors discussed models where the photon is composite; in particular, in this work we discuss a possibility that the photon is a composite state of Dirac singletons (see a discussion in the next section). An indirect confirmation of our conclusions is that all known neutral particles are bosons.

Another consequence of the fact that the IRs are implemented on the both hyperboloids is that there is no superselection rule prohibiting states which are superpositions of a particle and its antiparticle, and transitions particle$\leftrightarrow$antiparticle are not prohibited. As a result, the electric charge and the baryon and lepton quantum numbers can be only approximately conserved. In particular, they are approximately conserved if Poincare approximation works with a high accuracy.

This shows that dS invariant theory implies a considerably new understanding of the notion of particles and antiparticles. In contrast to Poincare or AdS theories, for combining a particle and its antiparticle together, there is no need to construct a local covariant field since they are already combined at the level of IRs.

This is an important argument in favor of dS symmetry. Indeed, the fact that in AdS and Poincare invariant theories a particle and its antiparticle are described by different IRs means that they are different objects. Then a problem arises why they have the same masses and spins but opposite charges. In QFT this follows from the CPT theorem which is a consequence of locality since *we construct* local covariant fields from a particle and its antiparticle with equal masses. A question arises what happens if locality is only an approximation: in that case the equality of masses, spins *etc.*, is exact or approximate? Consider a simple model when electromagnetic and weak interactions are absent. Then the fact that the proton and the neutron have the same masses and spins has nothing to do with locality; it is only a consequence of the fact that the proton and the neutron belong to the same isotopic multiplet. In other words, they are simply different states of the same object—the nucleon. We see, that in dS invariant theories the situation is analogous. The fact that a particle and its antiparticle have the same masses and spins but opposite charges (in the approximation when the notions of particles, antiparticles and charges are valid) has nothing to do with locality or non-locality and is simply a consequence of the fact that they are different states of the same object since they belong to the same IR.

The non-conservation of the baryon and lepton quantum numbers has been already considered in models of Grand Unification but the electric charge has been always believed to be a strictly conserved quantum number. In our approach all those quantum numbers are not strictly conserved because in the case of dS symmetry transitions between a particle and its antiparticle are not prohibited. The experimental data that these quantum numbers are conserved reflect the fact that at present Poincare approximation works with a very high accuracy. As noted in Sec. 1.4, the cosmological constant is not a fundamental physical quantity and if the quantity R is very large now, there is no reason to think that it was large always. This completely changes the status of the problem known as "baryon asymmetry of the World" since at early stages of the World transitions between particles and antiparticles had a much greater probability.

One might say that a possibility that only fermions can be elementary is not attractive since such a possibility would imply that supersymmetry is not fundamental. There is no doubt that supersymmetry is a beautiful idea. On the other hand, one might say that there is no reason for nature to have both, elementary fermions and elementary bosons since the latter can be constructed from the former. A well-known historical analogy is that the simplest covariant equation is not the Klein-Gordon equation for spinless fields but the Dirac and Weyl equations for the spin 1/2 fields since the former is the equation of the second order while the latter are the equations of the first order.

In 2000, Clay Mathematics Institute announced seven Millennium Prize Problems. One of them is called "Yang-Mills and Mass Gap" and the official description of this problem can be found in Ref. [117]. In this description it is stated that the Yang-Mills theory should have three major properties where the first one is as follows: "It must have a "mass gap;" namely there must be some constant $\Delta > 0$ such that every excitation of the vacuum has energy at least Δ." The problem statement assumes that quantum Yang-Mills theory should be constructed in the framework of Poincare invariance. However, as follows from the above discussion, this invariance can be only approximate and dS invariance is more general. Meanwhile, in dS theory the mass gap does not exist. Therefore we believe that the problem has no solution.

10.5.2 Particle theory over a Galois field

In standard theory a difference between representations of the so(2,3) and so(1,4) algebras is that IRs of the so(2,3) algebra where the operators $M^{\mu 4}$ ($\mu = 0, 1, 2, 3$) are Hermitian can be treated as IRs of the so(1,4) algebra where these operators are anti-Hermitian and vice versa. Suppose now that one accepts arguments of Chap. 6 that fundamental quantum theory should be constructed over a Galois field rather than the field of complex numbers. As noted in Chap. 6, in GFQT a probabilistic interpretation is only approximate and hence Hermiticy can be only a good approximation in some situations. Therefore one cannot exclude a possibility that elementary particles can be described by modular analogs of IRs of the so(2,3) algebra while modular representations describing symmetry of macroscopic bodies are modular analogs of standard representations of the so(1,4) algebra. In view of this observation, in Chap. 8 we consider standard and modular IRs of the so(2,3) algebra in parallel in order to demonstrate common features and differences between standard and modular cases.

As noted in Chap. 6, GFQT does not contain infinities at all and all operators are automatically well defined. In my discussions with physicists, some of them commented this fact as follows. This is an approach where a cutoff (the characteristic p of the Galois field) is introduced from the beginning and for this reason there is nothing strange in the fact that the theory does not have infinities. It has a large number p instead and this number can be practically treated as infinite.

However, the difference between Galois fields and usual numbers is not only that the former are finite and the latter are infinite. If the set of usual numbers is visualized as a straight line from $-\infty$ to $+\infty$ then the simplest Galois field can be visualized not as a segment of this line but as a circumference (see Fig. 6.1 in Sec. 6.1). This reflects the fact that in Galois fields the rules of arithmetic are different and, as a result, GFQT has many unusual features which have no analogs in standard theory.

The Dirac vacuum energy problem discussed in Sec. 8.8 is a good illustration of this point. Indeed, in standard theory the vacuum energy is infinite and, if

GFQT is treated simply as a theory with a cutoff p, one would expect the vacuum energy to be of the order of p. However, since the rules of arithmetic in Galois fields are different from standard ones, the result of exact (i.e. non-perturbative) calculation of the vacuum energy is precisely zero.

The original motivation for investigating GFQT was as follows. Let us take standard QED in dS or AdS space, write the Hamiltonian and other operators in angular momentum basis and replace standard IRs for the electron, positron and photon by corresponding modular IRs. One might treat this motivation as an attempt to substantiate standard momentum regularizations (e.g., the Pauli-Villars regularization) at momenta p/R (where R is the radius of the World). In other terms this might be treated as introducing fundamental length of the order of R/p. We now discuss reasons explaining why this naive attempt fails.

One of the main results in Chap. 8 is that (see Sec. 8.2) *in GFQT the existence of antiparticles follows from the fact that any Galois field is finite. Moreover, the very existence of antiparticles might be an indication that nature is described rather by a finite field or ring than by complex numbers.* We believe that this result is not only very important but also extremely simple and beautiful. A simple explanation of the above result follows.

In standard theory a particle is described by a positive energy IR where the energy has the spectrum in the range $[mass, \infty)$. At the same time, the corresponding antiparticle is associated with a negative energy IR where the energy has the spectrum in the range $(-\infty, -mass]$. Consider now the construction of a modular IR for some particle. We again start from the rest state (where energy=mass) and gradually construct states with higher and higher energies. However, in such a way we are moving not along a straight line but along the circumference in Fig. 6.1. Then sooner or later we will arrive at the point where energy=-mass.

The fact that in GFQT a particle and its antiparticle belong to the same IR makes it possible to conclude that, in full analogy with the case of standard dS theory (see the preceding section), there are no neutral particles in the theory, the very notion of a particle and its antiparticle is only approximate and the electric charge and the baryon and lepton quantum numbers can be only approximately conserved. As shown in Sec. 8.7, if one tries to replace nonphysical annihilation and creation operators (a, a^*) by physical operators (b, b^*) related to antiparticles then the symmetry on quantum level is inevitably broken. In GFQT, by analogy with standard theory, it is possible not to introduce the notion of antiparticles but work by analogy with Dirac's hole theory. Then the symmetry on quantum level is preserved and, as shown in Sec. 8.8, in contrast to standard theory, the vacuum can be chosen such that the vacuum energy is not infinite but zero. This poses a problem whether there are physical reasons for such a choice of the vacuum.

As explained in Sec. 8.9, the spin-statistics theorem can be treated as a requirement that standard quantum theory should be based on complex numbers. This requirement also excludes the existence of neutral elementary particles.

Since GFQT can be treated as the modular version of both, dS and AdS standard theories, supersymmetry in GFQT is not prohibited. In Sec. 8.10 we discuss common features and differences between standard and modular IRs of the osp(1,4) algebra. One of the most interesting feature of the modular case is how supersymmetry describes Dirac singletons in GFQT. This question is discussed in the next section.

10.6 Dirac singletons

One might conclude that since in GFQT the photon cannot be elementary, this theory cannot be realistic and does not deserve attention. However, the nonexistence of neutral elementary particles in GFQT shows that the photon (and the graviton and the Higgs boson if they exist) should be considered on a deeper level. In Chap. 9 we argue that in GFQT a possibility that massless particles are composite states of Dirac singletons is even more attractive than in standard theory.

As it has been noted in Chap. 9, the seminal result by Flato and Fronsdal [113] poses a fundamental problem whether only Dirac singletons can be true elementary particles. In this case one has to answer the questions (see Sec. 9.1):

- a) Why singletons have not been observed yet.
- b) Why such massless particles as photons and others are stable and their decays into singletons have not been observed.

In the literature, a typical explanations of a) are that singletons are not observable because they cannot be considered in the Poincare limit or because in this limit the singleton four-momentum becomes zero or because the singleton field lives on the boundary of the AdS bulk or as a consequence of other reasons. As shown in Sec. 9.3, in standard theory semiclassical approximations for singletons in Poincare limit can be discussed in full analogy with the case of massive and massless particles. As a result, in the general case the energy of singletons in Poincare limit is not zero but, in contrast to the case of usual particles, singletons can have only two independent components of standard momentum, not three as usual particles. A problem arises whether such objects can be detected by standard devices, whether they have a coordinate description etc. At the same time, in standard theory there is no natural explanation of b).

While in standard theory there are four singleton IRs describing the Di and Rac singletons and their antiparticles, in GFQT only two IRs remain since standard Di and anti-Di now belong to the same IR and the same is true for standard Rac and anti-Rac. We use Di and Rac to call the corresponding modular IRs, respectively. Nevertheless, since each massless boson can be represented as a composite state of two Dis or two Racs, a problem remains of what representation (if any) is preferable. This problem has a natural solution if the theory is supersymmetric. Then the only

IR is the (modular) Dirac supermultiplet combining (modular) Di and (modular) Rac into one IR.

The main result of Chap. 9 is described in Sec. 9.6 where we explicitly describe a complete set of supersymmetric modular IRs taking part in the decomposition of the tensor product of two modular Dirac supersingleton IRs. In particular, by analogy with the Flato-Fronsdal result, each massless superparticle can be represented as a composite state of two Dirac supersingletons and one again can pose a question of whether only Dirac (super)singletons can be true elementary (super)particles.

This question is also natural in view of the following observation. As shown in Sec. 3.2, the dS mass m_{dS} and the standard Poincare mass m are related as $m_{dS} = Rm$ where R is the radius of the world, and, as shown in Sec. 9.3, the relation between the AdS and Poincare masses is analogous. If for example one assumes that R is of the order of $10^{26}m$ then the dS mass of the electron is of the order of 10^{39}. It is natural to think that a particle with such a dS mass cannot be elementary. Moreover, the present upper level for the photon mass is $10^{-18}ev$ which seems to be an extremely tiny quantity. However, the corresponding dS mass is of the order of 10^{15} and so even the mass which is treated as extremely small in Poincare invariant theory might be very large in de Sitter theories. Nevertheless, assuming that only (super)singletons can be true elementary (super)particles, one still has to answer the questions a) and b).

As explained in Sec. 8.3, a crucial difference between Dirac singletons in standard theory and GFQT follows. Since 1/2 in the Galois field is $(p+1)/2$, the eigenvalues of the operators h_1 and h_2 for singletons in GFQT are $(p+1)/2, (p+3)/2, (p+5)/2...$, i.e. huge numbers if p is huge. Hence the Poincare limit and semiclassical approximation for Dirac singletons in GFQT have no physical meaning and they cannot be observable. In addition, as noted in Chap. 6, the probabilistic interpretation for a particle can be meaningful only if the eigenvalues of all the operators M_{ab} are much less than p. Since for Dirac singletons this is not the case, their state vectors do not have a probabilistic interpretation. These facts give a natural answer to the question a).

For answering question b) we note the following. In standard theory the notion of binding energy (or mass deficit) means that if a state with the mass M is a bound state of two objects with the masses m_1 and m_2 then $M < m_1 + m_2$ and the quantity $|M - (m_1 + m_2)|c^2$ is called the binding energy. The binding energy is a measure of stability: the greater the binding energy is, the greater is the probability that the bound state will not decay into its components under the influence of external forces.

If a massless particle is a composite state of two Dirac singletons, and the eigenvalues of the operators h_1 and h_2 for the Dirac singletons in GFQT are $(p+1)/2, (p+3)/2, (p+5)/2...$ then, since in GFQT the eigenvalues of these operators should be taken modulo p, the corresponding eigenvalues for the massless particle are $1, 2, 3....$ Hence an analog of the binding energy for the operators h_1 and h_2 is p, i.e. a huge number. This phenomenon can take place only in GFQT: although, from the formal point of view, the Dirac singletons comprising the massless state do not interact with each other, the analog of the binding energy for the operators h_1 and h_2 is huge. In other words, the fact that all the quantities in GFQT are taken modulo p implies a very strong effective interactions between the singletons. It explains why the massless state does not decay into Dirac singletons and why free Dirac singletons effectively interact pairwise for creating their bound state.

As noted in the literature on singletons (see e.g. the review [115] and references therein), the possibility that only singletons are true elementary particles but they are not observable has some analogy with quarks. However, the analogy is not full. According to Quantum Chromodynamics, forces between quarks at large distances prevent quarks from being observable in free states. In GFQT Dirac singletons cannot be in free states even if there is no interaction between them; the effective interaction between Dirac singletons arises as a consequence of the fact that GFQT is based on the arithmetic modulo p. In addition, quarks and gluons are used for describing only strongly interacting particles while in standard AdS theory and in GFQT quarks, gluons, leptons, photons, W and Z bosons can be constructed from Dirac singletons.

As noted at the end of Sec. 9.5, singleton physics can be directly generalized to the case of higher dimensions, and this fact has been indicated in the literature on singletons (see e.g. the review [115] and references therein).

The above discussion shows that singleton physics in GFQT is even more interesting than in standard theory.

10.7 Open problems

As we argue in Sec. 1.1, the main reason of the crisis in quantum physics is that nature, which is fundamentally discrete, is described by continuous mathematics. We also note that any ultimate quantum theory cannot be based on continuous mathematics even because, as follows from from Gödel's incompleteness theorems, that mathematics is not self-consistent.

One of the main results of this work is that gravity can be described as a pure kinematical manifestation of de Sitter symmetry over a Galois field. In this approach G is not fundamental but a quantity which can be calculated. In Sec. 1.5 we argue that the very notion of interaction cannot be fundamental and interaction constants can be treated only as phenomenological parameters. In particular, the Planck length has no fundamental meaning and the notions of gravitational fields and gravitons are not needed.

In view of these results the following problems arise. Since gravity can be tested only on macroscopic level, any quantum theory of gravity should solve the problem of constructing position operator on that level. As noted in Secs. 5.8 and 10.3, in the literature this problem is not discussed because it is tacitly assumed that the position operator on quantum level is the same as in standard quantum theory, but this is a great extrapolation. In quantum theory it is postulated that any physical quantity is defined by an operator. However, quantum theory does not define explicitly how the operator corresponding to a physical quantity is related to the measurement of this quantity. As shown in Chap. 5, the mass operator for all known gravitational phenomena is fully defined by a function describing the classical distance between the bodies in terms of their relative wave function. Therefore a fundamental problem is to understand the physical meaning of parameters characterizing wave functions of macroscopic bodies.

In our approach quantum theory is based on a Galois field with the characteristic p. A problem arises whether p is a constant or it is different in different periods of time. Moreover, in view of the problem of time in quantum theory, an extremely interesting scenario is that the world time is defined by p. As shown in Chap. 5, gravity is defined by the width of the distribution of the relative dS momentum. As argued in Sec. 7.2, the width depends on p as lnp and the gravitational constant in dS theory depends on p as $1/lnp$. Therefore the observable dynamics and what is treated as interactions might be simply manifestations of the fact that physics of our world depends on p.

As shown in Chap. 8, in our approach the notion of particle-antiparticle can be only approximate and the electric charge and other additive quantum numbers (e.g. the baryon and lepton quantum numbers) can be only approximately conserved. The extent of conservation depends on p: the greater is p, the greater is the extent of conservation. One might think that at present the conservation laws work with a high accuracy because the present value of p is extremely large. However, if at early stages of the world the value of p was much less than now then the conservation laws were not so strict as now. In particular, this might be a reason of the baryonic asymmetry of the world.

By analogy with gravity, one might think that electromagnetic, weak and strong interactions are not interactions but manifestations of higher symmetries. Similar ideas have been already extensively discussed in the literature, e.g. in view of compactification of extra dimensions.

The arguments and the results of this work give grounds to believe that sooner or later fundamental quantum theory will be discrete and finite and so it will be based either on finite fields or even on finite rings. In the present work we worked with a Galois field because working with a field is convenient from the technical point of view: in linear spaces over a field one can use the notions of basis and dimension. However, in view of the discussion in Chap. 6, division does not play a fundamental role in quantum theory and therefore a very interesting possibility is that the future quantum theory will involve only finite rings (this possibility has been pointed out by Metod Saniga).

Our results indicate that fundamental quantum theory has a very long way ahead (in agreement with Weinberg's opinion [118] that a new theory may be "centuries away").

Acknowledgements

I have greatly benefited from discussions with many physicists and mathematicians and it is difficult to mention all of them. A collaboration with Leonid Avksent'evich Kondratyuk and discussions with Skiff Nikolaevich Sokolov were very important for my understanding of basics of quantum theory. They explained that the theory should not necessarily be based on a local Lagrangian and symmetry on quantum level means that proper commutation relations are satisfied. Also, Skiff Nikolaevich told me about an idea that gravity might be a direct interaction. Eduard Mirmovich has proposed an idea that only angular momenta are fundamental physical quantities [119]. This has encouraged me to study de Sitter invariant theories. At that stage the excellent book by Mensky [46] was very helpful. I am also grateful to Sergey Dolgobrodov, Boris Hikin, Volodya Netchitailo, Mikhail Aronovich Olshanetsky, Michael Partensky, Metod Saniga and Teodor Shtilkind for numerous useful discussions.

Bibliography

[1] P.A.M. Dirac, *The Mathematical Foundations of Quantum Theory*, pp. 1-8 in *Mathematical Foundations of Quantum Theory*, A.R. Marlow ed. Academic Press: New York (1978).

[2] S. Weinberg. *The Quantum Theory of Fields*, Vol. I, Cambridge University Press: Cambridge, UK (1999).

[3] L.D. Landau and E.M. Lifshits, *Field Theory*. Nauka: Moscow (1973).

[4] B.L. Hikin, *Tensor Potential Description of Matter and Space, III. Gravitation*. arXiv:0803.1693 (gr-qc) (2008).

[5] B.L. Hikin, *Units of a Metric Tensor and Physical Interpretation of the Gravitational Constant*. viXra.org e-Print archive, viXra:1012.0038 (2010).

[6] E. Verlinde, *On the Origin of Gravity and the Laws of Newton*. JHEP **2011**, Issue 4, Article 29 (2011).

[7] W. Pauli, *General Principles of Quantum Mechanics*. Springer-Verlag: Berlin (1980). The original German edition: Handbuch der Physik, Vol. 5, *Prinzipien der Quantentheorie*. Springer-Verlag: Berlin (1958).

[8] Y. Aharonov and D. Bohm, *Time in the Quantum Theory and the Uncertainty Relation for Time and Energy*. Phys. Rev., **122**, 1649-1658 (1961).

[9] C. Rovelli, *Forget Time*. The FQXI Essay Contest "The Nature of Time" (2008).

[10] G. Keaton, *What is Time?* The FQXI Essay Contest "The Nature of Time" (2008).

[11] W. Pauli, *The Connection Between Spin and Statistics*. Phys. Rev. **58**, 716-722 (1940).

[12] A.I. Akhiezer and V.B. Berestetsky, *Quantum Electrodynamics*. Nauka: Moscow (1969).

[13] N.N. Bogolubov, A.A. Logunov, A.I. Oksak and I.T. Todorov, *General Principles of Quantum Field Theory*. Nauka: Moscow (1987).

[14] R. Haag, *Local Quantum Physics*. Springer: Berlin (1996).

[15] C.Rovelli, *Loop Quantum Gravity: the First Twenty Five Years.*. Class. Quantum Gravity **28**, 153002 (2011).

[16] V.B. Berestetsky, E.M. Lifshits and L.P. Pitaevsky, *Relativistic Quantum Theory*. Vol. IV, Part 1. Nauka: Moscow (1968).

[17] F.M. Lev, *Positive Cosmological Constant and Quantum Theory*. Symmetry **2**(4), 1401-1436 (2010).

[18] F. Lev, *de Sitter Symmetry and Quantum Theory*. Phys. Rev. **D85**, 065003 (2012).

[19] P.A.M. Dirac, *Forms of Relativistic Rynamics*. Rev. Mod. Phys., **21**, 392-399 (1949).

[20] T.D. Newton and E.P. Wigner, *Localized States for Elementary Systems*. Rev. Mod. Phys., **21**, 400-405 (1949).

[21] E. Inonu and E.P. Wigner, *Representations of the Galilei Group*. Nuovo Cimento, **IX**, 705 (1952).

[22] E.P. Wigner, *On Unitary Representations of the Inhomogeneous Lorentz Group*. Ann. Math., **40**, 149-204 (1939).

[23] S. Perlmutter, G. Aldering, G. Goldhaber, R.A. Knop, P. Nugent, P.G. Castro, S. Deustua, S. Fabbro, A. Goobar, D.E. Groom et. al., *Measurement of Omega and Lambda from H42 High-redshift Supernovae*. Astrophys. J., **517**, 565-586 (1999).

[24] A. Melchiorri, P.A.R. Ade, P. de Bernardis, J.J. Bock, J. Borrill, A. Boscaleri, B.P. Crill, G. De Troia, P. Farese, P.G. Ferreira et. al., *A Measurement of Omega from the North American Rest Flight of Boomerang*. Astrophys. J. **536**, L63-L66 (2000).

[25] D.N. Spergel, R. Bean, O. Dore, M.R. Nolta, C.L. Bennett, J. Dunkley, G. Hinshaw, N. Jarosik, E. Komatsu, L. Page et. al., *Wilkinson Microwave Anisotropy Probe (WMAP) Three Year Results: Implications for Cosmology*. Astrophys. J. Suppl., **170**, 377-408 (2007).

[26] K. Nakamura and Particle Data Group. J. Phys., **G37**, 075021 (2010) - see section "The Cosmological Parameters", which also can be found at http://pdg.lbl.gov/2010/reviews/rpp2010-rev-cosmological-parameters.pdf .

[27] J. B. Hartle, S. W. Hawking and T. Hertog, *Accelerated Expansion from Negative* Λ. arXiv:1205.3807 (2012).

[28] M.J. Duff, L.B. Okun and G. Veneziano, *Trialogue on the Number of Fundamental Constants*. JHEP, **3**, 023 (2002).

[29] L.B. Okun, *The "Relativistic" Mug*. arXiv:1010.5400 (gen-ph).

[30] S. Weinberg, H.B. Nielsen and J.G. Taylor, *Overview of Theoretical Prospects for Understanding the Values of Fundamental Constants*. In *The Constants of Physics*. W.H. McCrea and M.J. Rees eds. Phil. Trans. R. Soc. London, **A310**, 249-252 (1983).

[31] E.T. Akhmedov, A. Roura and A. Sadofyev, *Classical Radiation by Free-falling Charges in de Sitter Spacetime*. Phys. Rev., **D82**, 044035 (2010); E.T. Akhmedov, and P.A. Burda, *Simple Way to Take into Account Back Reaction on Pair Creation*. Phys. Lett., **B687**, 267-270 (2010); E.T. Akhmedov, *Real or Imaginary? (On Pair Creation in de Sitter Space)* Mod. Phys. Lett. **A25**, 2815-2823 (2010).

[32] E. Bianchi and C. Rovelli. *Why all These Prejudices Against a Constant?* arXiv:1002.3966v3 (astro-ph.CO) (2010).

[33] A Liddle, *An Introduction to Modern Cosmology* (2nd ed.). John Wiley & Sons Ltd.: London (2003).

[34] V.S. Netchitailo, *World-Universe Model*. viXra:1303.0077 (2013).

[35] F. Lev, *Finiteness of Physics and its Possible Consequences*. J. Math. Phys., **34**, 490-527 (1993).

[36] F.M. Lev, *Some Group-theoretical Aspects of SO(1,4)-Invariant Theory*. J. Phys. **A21**, 599-615 (1988); *The Problem of Interactions in de Sitter Invariant Theories*. J. Phys., **A32**, 1225-1239 (1999).

[37] F.M. Lev, *Could Only Fermions Be Elementary?* J. Phys., **A37**, 3287-3304 (2004).

[38] N.D. Birrell and P.C.W. Davies, *Quantum Fields in Curved Space*. Cambridge University Press: Cambridge, Great Britain (1982).

[39] R.M. Wald, *Quantum Field Theory in Curved Space-Time and Black Hole Thermodynamics*. The University of Chicago Press: Chicago (1995).

[40] F. Lev, *Representations of the de Sitter Algebra Over a Finite Field and Their Possible Physical Interpretation*. Yad. Fiz., **48**, 903-912 (1988); *Modular Representations as a Possible Basis of Finite Physics*. J. Math. Phys., **30**, 1985-1998 (1989); *Quantum Theory on a Galois Field*. hep-th/0403231 (2004).

[41] F. Lev, *Massless Elementary Particles in a Quantum Theory over a Galois Field*. Theor. Math. Phys. **138**, 208-225 (2004).

[42] F. Lev, *Why is Quantum Theory Based on Complex Numbers?* Finite Fields and Their Applications, **12**, 336-356 (2006).

[43] F.M. Lev, *Introduction to a Quantum Theory over a Galois Field*. Symmetry **2**(4), 1810-1845 (2010).

[44] F.M. Lev, *A Possible Mechanism of Gravity. hep-th/0307087 (2003)*; *De Sitter Invariance and a Possible Mechanism of Gravity*. arXiv:0807.3176 (gr-qc) (2008).

[45] F.M. Lev, *Is Gravity an Interaction?* Physics Essays, **23**, 355-362 (2010).

[46] M.B. Mensky, *Method of Induced Representations. Space-time and Concept of Particles*. Nauka: Moscow (1976).

[47] L.D. Landau and E.M. Lifshits, *Quantum Mechanics. Nonrelativistic Theory*. Nauka: Moscow (1974); L.E. Ballentine, *Quantum Mechanics. A Modern Development*. World Scientific: Singapore (2010).

[48] E. Schrödinger, *Collected Papers on Wave Mechanics*. Translated from the second German edition of the author's book *Abhandlungen zur Wellenmechanik*. Blackie & Son Limited: London and Glasgow (1929).

[49] P.A.M. Dirac, *The Principles of Quantum Mechanics*. Oxford University Press: Oxford (1982); L.I. Schiff, *Quantum mechanics*. McGraw-Hill: London (1968); E. Abers, *Quantum mechanics*. Pearson Education Inc.: Upper Saddle River, New Jersey (2004).

[50] A.O. Barut, *Quantum Theory of Single Events: Localized de Broglie Wavelets, Schrodinger Waves and Classical Trajectories*. Foundations of Physics, **20**, 1223-1240 (1990); A.O. Barut, *Coherent states versus de Broglie wavelets*. Proceedings of Conference: C93-06-12.1 IC-93-233, C93-06-12.1, Aug 1993 (ICTP, Trieste).

[51] L.G. Sapogin, *An Unitary Unified Quantum Field Theory*. Global Journal of Science Frontier Research, **11**, 47-74 (2011).

[52] S.J. Plimpton and W.E. Lawton, *A Very Accurate Test of Coulomb's Law of Force Between Charges*. Phys. Rev. **50**, 1066-1072 (1936).

[53] C.G. Darwin, *Free Motion in the Wave Mechanics*. Proc. R. Soc. London, **A117**, 258-293 (1927).

[54] M.V. Berry and N.L. Balazs, *Nonspreading Wave Packets*. Am. J. Phys. **47**, 264-267 (1979).

[55] G.A. Siviloglou, J. Broky, A. Dogariu, and D.N. Christodoulides, *Observation of Accelerating Airy Beams*. Phys. Rev. Lett. **99**, 213901 (2007).

[56] F. Lev, *Exact Construction of the Electromagnetic Current Operator in Relativistic Quantum Mechanics*. Ann. Phys. **237**, 355-419 (1995).

[57] G.C. Hegerfeldt, *Instantaneous Spreading and Einstein Causality in Quantum Theory*. Annalen Phys. **7**, 716-725 (1998).

[58] R.B. Griffiths, *Nonexistence of Quantum Nonlocality*. arXiv:1304.4425 (2013).

[59] M. Hawton, *Photon Wave Mechanics and Position Eigenvectors*. Phys. Rev. **A75**, 062107 (2007); *Photon Position Measure*. Phys. Rev. **A82**, 012117 (2010); *Photon Location in Spacetime*. Phys. Scr. **T147**, 014014 (2012).

[60] M. H. L.Pryce, *The Mass-Centre in the Restricted Theory of Relativity and Its Connexion with the Quantum Theory of Elementary Particles*. Proc. R. Soc. London **A195**, 62-81 (1948).

[61] G. Dillon, *Fourier Optics and Time Evolution of De Broglie Wave Packets*. arXiv:1112.1242 (2011).

[62] L. Mandel and E. Wolf, *Optical Coherence and Quantum Optics*. Cambridge University Press: New York (1995); M.O. Scully and M.S. Zubairy, *Quantum Optics*. Cambridge University Press: New York (1997); W.P. Schleich, *Quantum Optics in Phase Space*. Wiley-VCH Verlag: Berlin (2001).

[63] L.D. Landau and L.M. Lifshits, *Statistical Physics. Part 1*. Nauka: Moscow (1976).

[64] NASA homepage on gamma-ray bursts. http://imagine.gsfc.nasa.gov/docs/science/know_l1/bursts.html.

[65] A. Challinor, *CMB anisotropy science: a review*. Invited science review for Proceedings of IAU Symposium 288, "Astrophysics from Antarctica". arXiv:1210.6008 (2012).

[66] CERN homepage on the lead-proton-run. http://www.stfc.ac.uk/resources/PDF/UKnewsfromCERNIssue12FINAL.pdf

[67] Cassini Solstice Mission. NASA homepage for the Cassini mission and Huygens Titan probe. http://saturn.jpl.nasa.gov/.

[68] I.I. Shapiro, G.H. Pettengill, M.E. Ash, M.L. Stone, W.B. Smith, R.P. Ingalls, and R.A. Brockelman, *Fourth Test of General Relativity: Preliminary Results*. Phys. Rev. Lett. **20**, 1265-1269 (1968).

[69] http://www.haystack.mit.edu/hay/history.html. The official site of the Haystack radio telescope.

[70] http://deepspace.jpl.nasa.gov/dsn/antennas/34m.html. The official site of the DSS-25 antenna.

[71] B. Bertotti, L. Iess and P. Tortora, *A Test of General Relativity Using Radio Links with the Cassini Spacecraft*. Nature **425**, 374-376 (2003).

[72] A. Hewish, S.J. Bell, J.D.H. Pilkington, P.F. Scott and R.A. Collins, *Observation of a Rapidly Pulsating Radio Source*. Nature **217**, 709-713 (1968).

[73] V. Letokhov and S. Johansson, *Astrophysical Lasers*. Oxford University Press Inc.: New York (2009).

[74] K.Z. Hatsagortsyan and G.Yu. Kryuchkyan, *Photon-photon Interaction in Structured QED Vacuum*. International Journal of Modern Physics: Conference Series **15**, 22-30 (2012).

[75] C.I. Christov, *On the Evolution of Localized Wave Packets Governed by a Dissipative Wave Equation*. Wave Motion **45**, 154-161 (2008).

[76] K. Wang and Z. Cao, *Wave Packet for Massless Fermions and its Implication to the Superluminal Velocity Statistics of Neutrino*. arXiv:1201.1341 (2012).

[77] F. Lev, *Do We Need Dark Energy to Explain the Cosmological Acceleration?* J. Mod. Phys. **9A**, 1185-1189 (2012).

[78] H.S. Snyder, *Quantized Space-Time*. Phys. Rev. **71**, 38-41 (1947).

[79] L. Smolin, *Classical paradoxes of locality and their possible quantum resolutions in deformed special relativity*. General Relativity and Gravitation **43**, 3671-3691 (2011).

[80] N.T. Evans, *Discrete Series for the Universal Covering Group of the 3+2 de Sitter Group*. J. Math. Phys., **8**, 170-184 (1967).

[81] E.W. Mielke, *Quantenfeldtheorie im de Sitter-Raum*. Fortschr. Phys. **25**, 401-457 (1977).

[82] G.W. Mackey, *Induced Representatios of Locally Compact Groups*. Ann. Math. **55**, 101-139 (1952); **58**, 193-221 (1953); M.A. Naimark, *Normalized Rings*. Nauka: Moscow (1968); J. Dixmier, *von Neumann Algebras*. North-Holland Mathematical Library, Vol. 27. North-Holland Publishing Co.: Amsterdam-New York-Oxford (1977); A.O. Barut and R. Raczka, *Theory of Group Representations and Applications*. Polish Scientific Publishers: Warsaw (1977).

[83] N. Ikeda and T. Fukuyuama, *Fermions in (Anti) de Sitter Gravity in Four Dimensions*. Prog. Theor. Phys. **122**, 339-353 (2009).

[84] B. Braden, *Restricted Representatins of Classical Lie Algebras of Types A_2 and B_2*. Bull. Amer. Math. Soc. **73**, 482-486 (1967).

[85] L.A. Kondratyuk and M.V. Terent'ev, *The Scattering Problem for Relativistic Systems with a Fixed Number of Particles in Light-front Dynamics*. Sov. J. Nucl. Phys. **31**, 1087-1117 (1980).

[86] M.G. Fuda, *Poincare Invariant Potential Model*. Phys. Rev. **C36**, 1489-1506 (1987); *A New Picture For Light Front Dynamics*. Ann. Phys. **197**, 265-299 (1990).

[87] W.B. Hughes, *SU(2)×SU(2) Shift Operators and Representations of SO(5)*. J. Math. Phys. **24**, 1015-1020 (1983).

[88] A.L. Fetter and J.D. Walecka, *Quantum Theory of Many-Particle Systems*. Dover Publications Inc.: Mineola N.Y. (2003).

[89] D. Giulini, *Superselection Rules*. arXiv:0710.1516v2 (quant-ph) (2007).

[90] E.T. Akhmedov and P.V. Buividovich, *Interacting Field Theories in de Sitter Space are Non-Unitary*. Phys. Rev. **D78**, 104005 (2008).

[91] H. Bateman and A. Erdelyi, *Higher Transcendental Functions*. Mc Graw-Hill Book Company: New York (1953).

[92] L.B. Okun, *Photons and Static Gravity*. Mod. Phys. Lett. **A15**, 1941-1947 (2000).

[93] S. Weinberg, *Gravitation and Cosmology: Principles and Applications of the General Theory of Relativity*. John Wiley & Sons, Inc. New York - London - Sydney - Toronto (1972).

[94] M.D. Scadron, *Advanced Quantum Theory*. World Scientific, London - Hackensack, NJ (2007).

[95] J.M. Weisberg and J.H. Taylor, *The Relativistic Binary Pulsar B1913+16: Thirty Years of Observations and Analysis*, in *Proc. Aspen Conference*, Astronomical Society of the Pacific Conference Series, F.A. Rasio and I.H. Stairs eds. **328** (2005).

[96] J. M. Weisberg, D.J. Nice and J. H. Taylor, *Timing Measurements of the Relativistic Binary Pulsar PSR B1913+16*. Astrophysical J., **722**, 1030-1034 (2010).

[97] C.M. Will, *The Confrontation between General Relativity and Experiment.* Living Reviews in Relativity. **9**, Irr-2006-3, (2006).

[98] B.L. Van der Waerden, *Algebra I*. Springer-Verlag: Berlin - Heidelberg - New York (1967); K. Ireland and M. Rosen, *A Classical Introduction to Modern Number Theory* (Graduate Texts in Mathematics-87. Springer: New York - Heidelberg - Berlin (1987); H. Davenport, *The Higher Arithmetic*. Cambridge University Press: Cambridge (1999).

[99] S. Weinberg, *Living with Infinities*. arXiv:0903.0568 (hep-th)(2009).

[100] M. Planat and M. Saniga, *Finite Geometries in Quantum Theory: From Galois (fields) to Hjelmslev (rings)*. J. Mod. Phys., **B20**, 1885-1892 (2006).

[101] H.R. Coish, *Elementary Particles in a Finite World Geometry*. Phys. Rev. **114**, 383-388 (1959); I.S. Shapiro, *Weak Interactions in the Theory of Elementary Particles with Finite Space*. Nucl. Phys. **21**, 474-491 (1960); Y. Nambu, *Field Theory of Galois Fields*, pp. 625-636 in *Quantum Field Theory and Quantum Statistics.* I.A. Batalin et. al. eds. Adam Hilger: Bristol (1987); A. Vourdas, *Quantum Systems with Finite Hilbert Space*. Rep. Progr. Phys. **67**, 267-320 (2004).

[102] L.N Chang, Z. Lewis, D. Minic and T. Takeuchi, *Quantum systems based upon Galois fields from sub-quantum to super-quantum correlations.* Int. J. Mod. Phys. **A29**, 1430006 (2014).

[103] S.L. Adler, *Quaternionic Quantum Mechanics and Quantum Fields*. Oxford University Press: Oxford England (1995); B. Dragovich, A. Yu. Khrennikov, S. V. Kozyrev, and I. V. Volovich, *On p-adic Mathematical Physics*. p-Adic Numbers, Ultrametric Analysis and Applications **1**, 1-17 (2009).

[104] J.C. Jantzen, *Representations of Lie algebras in prime characteristic*. NATO ASI series, series C, Math. Phys. Sci. "Representation Theories and Algebraic Geometry". **514**, 185-235 (1998).

[105] H. Zassenhaus, *The Representations of Lie Algebras of Prime Characteristic.* Proc. Glasgow Math. Assoc., **2**, 1-36 (1954).

[106] R. Steinberg, *Lectures on Chevalley Groups*. Yale University: New Haven, CT (1967).

[107] F.M. Lev, *Supersymmetry in a Quantum Theory over a Galois Field*. hep-th/0209221 (2002).

[108] P. Billingsley, *Probability and Measure*. Wiley Series in Probability and Mathematical Statistics: New York (1995).

[109] P.A.M. Dirac, *A Remarkable Representation of the 3 + 2 de Sitter group*. J. Math. Phys. **4**, 901-909. (1963).

[110] S. Weinberg, *The Quantum Theory of Fields, Volume III Supersymmetry*. Cambridge University Press: Cambridge, United Kingdom (2000).

[111] W. Heidenreich, *All Linear Unitary Irreducible Representations of de Sitter Supersymmetry with Positive Energy*. Phys. Lett. **B110**, 461-464 (1982).

[112] C. Fronsdal, *Dirac Supermultiplet*. Phys. Rev. **D26**, 1988-1995 (1982).

[113] M. Flato and C. Fronsdal, *One Massles Particle Equals two Dirac Singletons*. Lett. Math. Phys. **2**, 421-426 (1978).

[114] W. Heidenreich, *Tensor Product of Positive Energy Representations of* $\tilde{S}(3,2)$ *and* $\tilde{S}(4,2)$. J. Math. Phys. **22**, 1566-1574 (1981).

[115] M. Flato, C. Fronsdal and D. Sternheimer, *Singleton Physics*. hep-th/9901043 (1999).

[116] X. Bekaert, *Singletons and Their Maximal Symmetry Algebras*. arXiv:1111.4554 (2011).

[117] A.Jaffe and E. Witten, *Quantum Yang-Mills Theory*. An Official Description of One of Seven Clay Mathematics Institute Millenium Prize Problems. www.claymath.org/millennium/ .

[118] S. Weinberg, *Dreams of a Final Theory*. A Division of Random House Inc.: New York (1992).

[119] F.M. Lev and E.G. Mirmovich, *Aspects of de Sitter Invariant Theories*. VINITI No 6099 Dep. (1984).