Relativistic Quantum Mechanics

Janos Polonyi
University of Strasbourg
(Dated: April 25, 2019)

Contents

Quantum mechanics is presented on different levels at the regular university lectures, and these notes are supposed to bridge Quantum Mechanics II, a detailed discussion of the non-relativistic one- or two-body problems and Quantum Mechanics III, an introduction to second quantization. The lecturers usually rightly devote this latter to the technical complexity of the subject which requires indeed time and attention. But this leaves the students somehow loosing the continuity and the familiarity of the territory as they suddenly find themselves in the midst of formal quantum fields, Feynman graphs and scattering amplitudes after having trained themselves with calculating expectation values using down the Earth coordinate and momentum operators.

I believe that more elaboration is needed to prepare the formalism to deal with relativistic effects and to shift the attention from expectation values to transition amplitudes. These notes should help to fill up the gap as far as the first subject is concerned.

I. RELATIVISTIC CLASSICAL DYNAMICS

Special relativity is about the preservation of the laws of classical physical in different coordinate systems, called inertial reference frames. Rather than checking the fundamental equations one by one it is required that

1. a free point particle moves with constant speed and

2. the propagation of light takes place with the same velocity, c

in each reference frame. The first condition is used to find the Lorentz transformation,

$$x^\mu = \Lambda^\mu_{\ \nu} x^\nu, \tag{1}$$

relating the space-time coordinates, $x^\mu = (ct, \boldsymbol{x})$, of different reference frames. The classical and the quantum equations of motion transform in a covariant manner during these transformation. The second condition is satisfied by requiring that the Lorentz transformations preserve

$$s^2 = t^2 - \boldsymbol{x}^2 = x^\mu g_{\mu\nu} x^\nu. \tag{2}$$

The family of linear transformations satisfying this conditions,

$$g_{\mu\nu} = \Lambda^{\mu'}_{\ \nu} g_{\mu'\nu'} \Lambda^{\mu'}_{\ \nu}, \tag{3}$$

defines the Lorentz group.

The motion of a point particle is described by its world-line, $x^\mu(s)$, s being a parameter which can be chosen as the invariant length in case of a massive particle. The invariant length, s, is called proper time since it gives the time shown by a clock, attached to the particle as long as it moves in the absence of external force.

The importance of condition 2 is the it introduces the same internal velocity parameter in the dynamics for all particles. This generates two velocity regimes, $v/c \sim 0$ and $v/c \sim 1$. The Newtonian mechanics, $v/c \to 0$, has no internal scale.

A. Anti-particles

The space-time is not an à priori structure in Special Relativity, we make it up with the help of meter rods and clocks. Note that this procedure is limited to the macroscopic, classical regime, for there is no way to measure the coordinates within an atom. The motion of a point particle is described by means of a trajectory, $\boldsymbol{x}(t)$, in non-relativistic mechanics and the time t plays a double role: It is a basic parameter of the dynamics and it orders events into a causal order, earlier phenomena assumed to be the origin of later ones. The time is a fundamentally different variable than the coordinate and this remains so in Special Relativity but the roles, mentioned above, are taken over by another variable, the parameter of the world line.

Do the non-relativistic trajectory, $\boldsymbol{x}(t)$ and the relativistic world line, $x^\mu(x)$, cover the same possibilities in physics? It is clear that one can find a world line for each trajectory but the converse is not true. The relation between the coordinate system and the proper time,

$$ds^2 = dt^2 \left(1 - \frac{\boldsymbol{v}^2}{c^2}\right), \tag{4}$$

allows the time and the proper time run either parallel or in opposite directions,

$$ds = \pm dt \sqrt{1 - \frac{\boldsymbol{v}^2}{c^2}}, \tag{5}$$

the latter having no place in non-relativistic physics. In other words, the coordinate time may go backward compared to the flow of time of a particle for a part of the world line as shown in Fig. 1. The particle with opposite time flow is called anti-particle and it obviously possesses the same mass as the particle. The world line of Fig. 1 describes a particle-anti particle pair beside the original particles for $t_1 < t < t_2$. The energy-momentum conservation is maintained at $t = t_1$ and t_2 by absorbing and emitting, respectively another particle, indicated by dashed lines in the figure.

Note that the world-line must have a sudden change of four velocity, $\dot{x}^\mu = dx^\mu/ds$, when the direction of time flips because the velocity, $\boldsymbol{v} = d\boldsymbol{x}/dt$, is bounded by $c(= 1)$. The discontinuous

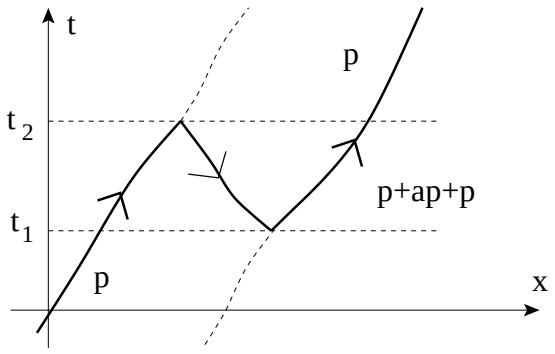

FIG. 1: A world line of a point particle which does not exists in non-relativistic mechanics.

jump of the momentum of the particle relegates the physical processes, taking place at these points, beyond classical physics. The reason is that classical physics, the unique determination of the future by the initial conditions, is inconsistent with such singularity. It is well known that the necessary condition that the solution of the differential equation,

$$\dot{x}(t) = f(x,t), \tag{6}$$

be uniquely determined by the initial condition, $x(t_i) = x_i$ is the continuity of $\partial_x f(x,t)$. To see the need of the continuous derivative consider the differential equation $\dot{x} = g|x|^p$ with $0 < p < 1$. It is easy to check that the trajectory,

$$x(t) = \begin{cases} 0 & t < t_0, \\ [g(q-p)(t-t_0)]^{\frac{1}{1-p}} & t \geq t_0, \end{cases} \tag{7}$$

solves the equation for $x_i = 0$ and an arbitrary $t_0 > t_i$, there are different solutions which share the same initial condition. Hence it is highly non-trivial to recover the traditional, classical picture where the future is determined by the initial conditions. If the particle possesses electric charge then a photon is annihilated and created at t_1 and t_2, respectively according to quantum electrodynamics. These processes represent quantum fluctuations which are known to be non-deterministic.

What is not excluded by conservation laws and may take place in classical physics can happen with some probability in quantum mechanics. Hence a quantum system containing particles and anti-particles may not conserve the number of degrees of freedom. This latter is fixed and is given by the number of independent variables of the wave function hence unitarity is expected to be violated when the usual formalism of quantum mechanics is applied with relativistic kinematics.

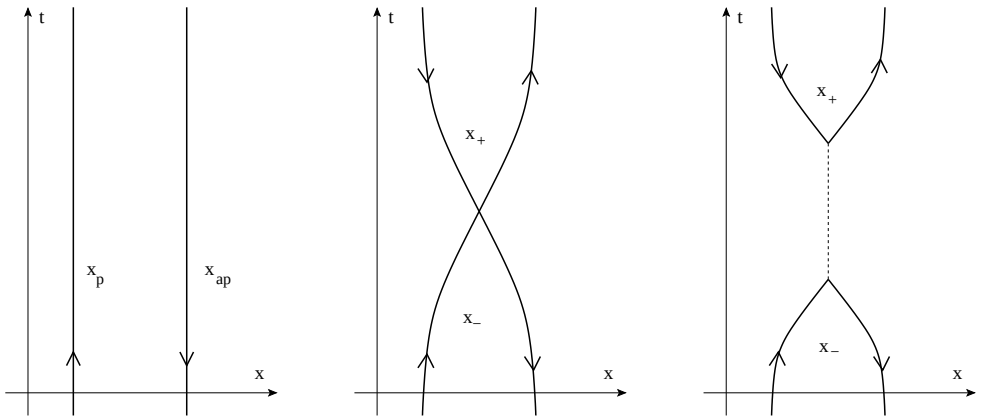

FIG. 2: (a): The world lines of a point-like particle, $x^\mu_p(s)$, and and an anti-particle, $x^\mu_{ap}(s)$, at rest in an 1+1 dimensional space-time. (b) The end points are changed and as a result the particle and anti-particle cross each other. We change the parametrization of the world lines to $x^\mu_-(s)$ and $x^\mu_+(s)$, covering the particle and anti-particles segments before and after the crossing, respectively. (c): The singularities of the world lines can be displaced in the space-time by continuous deformations and the particle-anti particle pair is first annihilated and later created.

B. Energy of an anti-particle

The invariance of the canonical equations of motion,

$$\dot{q} = \frac{\partial H(q,p)}{\partial p}, \quad \dot{p} = -\frac{\partial H(q,p)}{\partial q}, \tag{8}$$

remain invariant under the transformation $t \to -t$, $H \to -H$ for time independent Hamiltonian, H, suggesting that the energy of a particle and an anti-particle differ in their sign only, in agreement with the double-valued dispersion relation,

$$E = \pm c\sqrt{m^2c^2 + \boldsymbol{p}^2}, \tag{9}$$

of relativistic mechanics. This would cause serious difficulties in the quantum case. In fact, an important difference between the classical and quantum mechanics is that the boundedness of the Hamiltonian is necessary for a quantum particle. The reason is that an arbitrarily weak coupling to the fields, representing the environment, generates spontaneous decay processes towards the lower energy states. The emitted quanta is not recovered within reasonable time if the environment fills up sufficiently large volume and the resulting irreversibility generates singular, ill-defined time evolution if the particle Hamiltonian is unbounded from below. It is easy to see that a system of non-interacting bosons whose energy can be negative has no ground state. In fact, the energy is an additive observable and by adding a particle with negative energy one can lower the total energy.

Thus it is important to check if the energy of an anti-particle is indeed, without any doubt, negative. There are two ways the energy can be defined in classical physics, both are based on the action. Therefore let us start with the choice of the action of a relativistic particle. It should be Lorentz and translation invariant, the simplest choice is

$$S = -\frac{mc}{2}\int ds f(\dot{x}^\mu(s) g_{\mu\nu}\dot{x}^\mu(s)) \tag{10}$$

where s denotes the parameter of the world line of length dimension, $f(z)$ stands for a dimensionless function and the prefactor is introduced to have the desired dimension, that of the angular momentum. The mass m is rendered unique by requiring $f'(1) = 1$ and the sign is to assure the correct non-relativistic limit. The parameter s is not associated with the invariant length at this stage because such a relation would reduce the number of independent space-time coordinates in the variational calculus which would be an unwanted complication. Instead, s is identified with the invariant length after the derivation of the Euler-Lagrange equation,

$$\begin{aligned} 0 &= -mc\frac{d}{ds}f'(\dot{x}^2)\dot{x}^\mu \\ &= -mcf'(\dot{x}^2)\ddot{x}^\mu - 2mcf''(\dot{x}^2)\dot{x}_\nu\ddot{x}^\nu\dot{x}^\mu. \end{aligned} \tag{11}$$

If s is the invariant length then $\dot{x}^2 = 1$ and the derivative of this identity with respect to s, $\dot{x}_\nu\ddot{x}^\nu = 0$, simplifies the equation of motion to the expected form,

$$0 = mc\ddot{x}^\mu. \tag{12}$$

A possible definition of the energy comes form the canonical energy-momentum,

$$p_\mu = -\frac{\partial L}{\partial \dot{x}^\mu} = mc\dot{x}_\mu, \tag{13}$$

where the minus sign is to have to correct non-relativistic limit for the spatial components of $p_\mu = (\frac{E}{c}, -\boldsymbol{p})$. The transformation $s \to -s$ change the sign of the energy. Another definition,

$$P_\mu = \int d^3x\sqrt{-g}T_{\mu 0}, \tag{14}$$

is based on the energy-momentum tensor, the source of the gravitational interaction,

$$T_{\mu\nu}(x) = \frac{2}{\sqrt{-g}}\frac{\delta S}{\delta g^{\mu\nu}(x)}, \tag{15}$$

where the action is rewriten in terms of a curvlinear coordinate system by the introduction of a metric tensor $g^{\mu\nu}(x)$. This expression arises from the derivation of Einnstein's equation of General

Relativity by varying the Einstein-Hilbert action with respect to $g^{\mu\nu}$. The action (10) contains the covariant metrix tensor whose variation can be found by varying the identity $\delta^\mu_\rho = g^{\mu\nu}g_{\nu\rho}$,

$$0 = \delta g^{\mu\nu} g_{\nu\rho} + g^{\mu\nu}\delta g_{\nu\rho}, \tag{16}$$

yielding

$$\delta g_{\mu\nu} = -g_{\mu\rho}g_{\nu\sigma}\delta g^{\rho\sigma}. \tag{17}$$

The action (10), rewritten for general coordinate system reads as

$$S = -\frac{mc}{2}\int dx ds \delta^{(4)}(x - x(s)) f(\dot{x}^\mu g_{\mu\nu}(x)\dot{x}^\nu) \tag{18}$$

which together with (17) gives the energy-momentum tensor

$$T^{\mu\nu}(x) = mc\int ds \delta^{(4)}(x - x(s))\dot{x}^\mu(s)\dot{x}^\nu(s). \tag{19}$$

The corresponding energy-momentum vector,

$$\begin{aligned} P^\mu &= m\int ds\delta(t - x^0(s))\dot{x}^\mu(s)\dot{x}^0(s) \\ &= m\dot{x}^\mu(s)\frac{\dot{x}^0(s)}{|\dot{x}^0(s)|}, \end{aligned} \tag{20}$$

differs from (13) in a sign, rendering the energy-momentum of a particle and an anti-particle identical.

We assume now that the particle carries an electric charge and moves in the presence of the electromagnetic field $A_\mu(x)$. The Lagrangian is chosen to be

$$L = -\frac{m}{2}\dot{x}^2 - e\dot{x}^\mu A_\mu(x), \tag{21}$$

as the simplest extension of the non-relativistic free particle Lagrangian which changes at most by a boundary term under gauge transformation, $A_\mu(x) \to A_\mu(x) + \partial_\mu\alpha(x)$, $\alpha(x)$ being an arbitrary function. We use here the choice $f(z) = z$ in the action (10) and e denotes the charge of the particle. The Lagrangian (21) looks like that of a charge moving in a magnetic field in 4 dimensional space, except that the signature of the metric is not definite.

C. Space-time inversions and charge conjugation

The elements of the Lorentz group, L, matrices satisfying (3) can be split into disconnected sets. By taking the determinant of eq. (3) we have $(\det\Lambda)^2 = 1$. Since the matrix Λ is real its

determinant is real, as well and we must have $\det\Lambda = \pm 1$. We define the subsets of the Lorentz group, L_+ and L_-, by collecting Lorentz matrices of determinant 1 and -1, respectively. the set of matrices L_+ and L_- can not be joined by a continuous path in the Lorentz group because the determinant is a continuous function of the matrix elements. Thus we have found two disconnected components of the Lorentz group, $L = L_+ \cup L_-$, of which L^+ forms a subgroup.

The matrix element (00) of eq. (3),

$$1 = (\Lambda^0{}_0)^2 - \sum_j (\Lambda^0{}_j)^2, \tag{22}$$

shows that

$$|\Lambda^0{}_0| \geq \sqrt{1 + \sum_j (\Lambda^0{}_j)^2} \geq 1, \tag{23}$$

hence elements of the sets $L^\uparrow$ and $L^\downarrow$, containing Lorentz transformation matrices with $\Lambda^0{}_0 \geq 1$ and $\Lambda^0{}_0 \leq -1$, respectively give again a partition of the Lorentz group into two disconnected sets, $L = L^\uparrow \cup L^\downarrow$, $L^\uparrow$ being a subgroup.

We have finally four disconnected components of the Lorentz group, $L_+^\uparrow = L_+ \cap L^\uparrow$, $L_+^\downarrow = L_+ \cap L^\downarrow$, $L_-^\uparrow = L_- \cap L^\uparrow$ and $L_-^\downarrow = L_- \cap L^\downarrow$. The connected subgroup, $L_+^\uparrow$, is called the proper Lorentz group. The there is a bijective map among the four components, realized by particular Lorentz transformations, by the space and the time inversions,

$$P: \ (t, \boldsymbol{x}) \to (t, -\boldsymbol{x}) \tag{24}$$

and

$$T: \ (t, \boldsymbol{x}) \to (-t, \boldsymbol{x}), \tag{25}$$

respectively, $L_+^\downarrow = PTL_+^\uparrow$, $L_-^\uparrow = PL_+^\uparrow$ and $L_-^\downarrow = TL_+^\uparrow$.

Lorentz transformations act on world lines, as well and they bring a physically realizable world line into another physical one. One can extend the symmetry group of the world lines by a further discrete transformation, $C : s \to -s$, by exploiting the sign ambiguity, (5), in relating the proper time and the coordinate time.

Note that the action, constructed by means of the Lagrangian (21) can be made invariant under the combined transformation CPT by extending the C transformation to the vector potential as a charge conjugation, $C : A \to -A$. This symmetry is not an accident, it is a rather non-trivial theorem of relativistic quantum field theory stating that all relativistically invariant, local

Lagrangian are CPT invariant. As a result, the classical Lagrangian, derived from a local quantum theory should respect this symmetry, as well.

It is advantageous introduce the parity of the discrete transformations, $P : f(x,\dot{x}) \to f(Px,P\dot{x}) = \pi_f f(x,\dot{x})$, $T : f(x,\dot{x}) \to f(Tx,T\dot{x}) = \tau_f f(x,\dot{x})$, $C : f(x,\dot{x}) \to f(x,-,\dot{x}) = \gamma_f f(x,\dot{x})$, eg. $\pi_t = -\pi_{\boldsymbol{x}} = 1$, $-\tau_t = \tau_{\boldsymbol{x}} = 1$ and $\gamma_{x^\mu} t = -\gamma_{\dot{x}^\mu} = 1$. Since any function $f(z)$ can be written as a sum of an even and an odd part,

$$f(z) = \frac{f(z)+f(-z)}{2} + \frac{f(z)-f(-z)}{2}, \tag{26}$$

these parities assume the values ± 1, in agreement with the relations

$$P^2 = T^2 = C^2 = \mathbb{1} \tag{27}$$

in classical physics.

These equations are not necessarily valid in relativistic quantum mechanics, as shown in Appendix A. The point is the existence of the spin, a relativistic effect. The coordinate space of a non-relativistic, spinless particle is R^3, a connected space. The spin makes it necessary to follow rotations, as well, and to extend the kinematical space over rotations. We shall see that the rotational group is multiply-connected, a feature of the kinematical space which leads to multi-valued wave functions. The relative phase between the Rieman-sheets of the wave function introduces new quantum numbers, changing the parities.

II. SCALAR PARTICLE

The relativistic generalization of the Schrödinger equation for a free particle,

$$i\hbar\partial_t \psi(\boldsymbol{x},t) = -\frac{\hbar^2}{2m}\Delta\psi(\boldsymbol{x},t), \tag{28}$$

should have the same order of derivative in time and space directions. This can be achieved in two different manners, both have either first or second order derivatives. The former and the latter equation, describe fermions and bosons, respectively.

A. Heuristic derivation of the Klein-Gordon equation

The heuristic argument to construct a second order equation of motion is based on the generalization of the three-momentum to

$$p_\mu = \left(\frac{E}{c}, -\boldsymbol{p}\right) = i\hbar\partial_\mu = i\hbar(\partial_0, \boldsymbol{\nabla}), \quad p^\mu = \left(\frac{E}{c}, \boldsymbol{p}\right) = i\hbar\partial^\mu = i\hbar(\partial_0, -\boldsymbol{\nabla}). \tag{29}$$

The mass shell condition,

$$p^2 = m^2c^2\dot{x}^2 = m^2c^2, \tag{30}$$

leads to the Klein-Gordon equation

$$\left(\Box + \frac{m^2c^2}{\hbar^2}\right)\phi(x) = 0 \tag{31}$$

as an equation of motion. The parameter $\hbar/mc = \lambda_C$, the Compton wavelength, is the intrinsic length scale of a massive particle. Most of the equations below are given in units $\hbar = c = 1$ for simplicity. The generalization for a charged particle in the presence of an electromagnetic field $A_\mu(x)$, $p^\mu \to p^\mu - \frac{q}{c}A^\mu$ is obtained by the help of the covariant derivative $\partial_\mu \to \partial_\mu + i\frac{q}{\hbar c}A_\mu = D_\mu$. Special care is neede to set the value of q, it stands for the charge of the field the covariant derivative is acting upon. Since the electric charge is defined as the Noether charge, corresponding to phase rotation, $\phi(x) \to e^{-i\frac{q}{\hbar c}\alpha}\phi(x)$ the fields $\phi(x)$ and $\phi^*(x)$ have opposite electric charge, e and $-e$, respectively and $D_\mu\phi = \partial_\mu\phi + i\frac{e}{\hbar c}A_\mu\phi$ and $D_\mu\phi^* = \partial_\mu\phi - i\frac{e}{\hbar c}A_\mu\phi = (D_\mu\phi)^*$. Further useful properties of the covariant derivative is that it satisfies Leibnitz's rule, $D\phi\chi = (D\phi)\chi + \phi D\chi$, and therefore allows partial integration. The generalization of (31) for vector potential reads as

$$\left(D_\mu D^\mu + \frac{m^2c^2}{\hbar^2}\right)\phi(x) = 0. \tag{32}$$

The Klein-Gordon equation can be considered as the Euler-Lagrange equation for a complex scalar field $\phi(x)$ when the action

$$S[\phi] = \int dtd^3xL(\phi, \phi^*, D_\mu\phi, D_\mu\phi^*), \tag{33}$$

defined by the Lagrangian

$$L = (D_\mu\phi)^* D^\mu\phi - m^2\phi^*\phi \tag{34}$$

is used. The Lagrangian is invariant under a continuous symmetry transformation, $\phi(x) \to e^{i\theta}\phi(x)$, and the corresponding conserved Noether-current is

$$j_\mu = \frac{i}{2m}(\phi^* D_\mu\phi - (D_\mu\phi)^*\phi) = \frac{i}{2m}\phi^*\overleftrightarrow{\partial_\mu}\phi - \frac{e}{mc}\phi^*\phi A_\mu, \tag{35}$$

where the derivative $\overleftrightarrow{\partial_\mu}$ is defined as

$$f\overleftrightarrow{\partial_\mu}g = g\partial_\mu g - \partial_\mu g f. \tag{36}$$

It is the relativistic generalization of the non-relativistic probability current

$$j_\mu = \left(\psi^*\psi, \frac{1}{2im}[\psi^*\boldsymbol{D}\psi - (\boldsymbol{D}\psi)^*\psi]\right) = \left(\psi^*\psi, \frac{1}{2im}(\psi^*\boldsymbol{\nabla}\psi - \boldsymbol{\nabla}\psi^*\psi) + \frac{e}{mc}\psi^*\psi\boldsymbol{A}\right) \tag{37}$$

of the Schrödinger equation. The $\mathcal{O}(A)$ terms in (35)-(37) represent the paramagnetic contribution to the electric current.

The plane-wave solutions are

$$\phi(x) = e^{\mp i p_\mu x^\mu} \tag{38}$$

where $p_0 = \omega_p$ with

$$\omega_p = \sqrt{m^2 + \boldsymbol{p}^2}. \tag{39}$$

The sign - or + corresponds to positive energy (particle) or negative energy (anti-particle) plane waves, respectively. The current (35) assumes the form

$$j^\mu = \pm \frac{p^\mu}{m} \tag{40}$$

for a plane wave. The negative energy plane wave represents anti-particles and the form (38) of the wave function suggests that the energy-momentum p^μ is the canonical one, corresponding to (13).

The Klein-Gordon equation leads to a number of paradoxes of which we mention two only:

1. *States with negative energy:* The energy of a free particle, defined by the eigenvalue of $i\partial_0$ is not definite. The appearance of negative energy one-particle states poses a serious problem in quantum mechanics. In fact, the energy of a system of several free bosons can be lowered without bound in this case and there is no ground state anymore.

2. *No probabilistic interpretation:* The density ρ of the Klein-Gordon current, (35) $j^\mu = (\rho, \boldsymbol{j})$, is the time component of a four vector rather than a scalar as in non-relativistic quantum mechanics. It can not be interpreted as probability density because signρ may change during Lorentz transformation.

Note that these problems are not independent, both of them arise from the same minus sign in the exponent of the plane wave (38), the existence of anti-particles.

B. First order formalism for scalar particles

The appearence of the anti-particle, the source of the sign problems, mentioned above, can easily be traced back to the increase of the order of the equation of motion in the time derivative. The point is that the unique solution of a second order differential equation, (32), needs two

initial conditions, for instance $\phi(t_i, \boldsymbol{x})$ and $\partial_0\phi(t_i, \boldsymbol{x})$. This is an important change with respect to non-relativistic quantum mechanics, suggesting the presence of two particles, one for each initial condition function, in the state, represented by $\phi(t, \boldsymbol{x})$. It is natural to identify the second particle with the anti-particle partner of the original particle. In other words, the wave function,

$$\phi(t, \boldsymbol{x}) = \int \frac{d\omega}{2\pi} e^{-i\omega t}\phi(\omega, \boldsymbol{x}), \tag{41}$$

is a linear superposition of the particle and anti-particle,

$$\phi^{\pm}(t, \boldsymbol{x}) = \Lambda_{\pm}\phi(t, \boldsymbol{x}) = \int dt' \Lambda_{\pm}(t - t')\phi(t', \boldsymbol{x}) \tag{42}$$

where the projectors

$$\begin{aligned} \Lambda_{\pm}(t - t') &= \pm \int_0^{\pm\infty} \frac{d\omega}{2\pi} e^{-i\omega(t-t' \mp i\epsilon)} \\ &= -\frac{i}{2\pi(t - t' \mp i\epsilon)}, \end{aligned} \tag{43}$$

separate the positive and negative frequency components therefore have to act on the time variable.

Owing to the unconstrained integration over t' in eq. (42) the separation of the particle and the anti-particle content of the wave function $\phi(t, \boldsymbol{x})$ is possible after infinitely long observation time. In fact, the energy-time uncertainty principle is due to the non-local nature of the Fourier transformation. This represents an important difference between the relativistic and the non-relativistic quantum mechanics, namely the separation of the two relativistic particles can not be done locally in time. The equation of motion can be used to make the separation locally, by relating time- and space-dependence for free particles. This solves the problem for free particles but any change in the equation of motion induces a mixing of the states, defined for the free particles . This is the way we understand the creation particle-anti particle pairs by external fields, examples being the electron-positron emission by strong electromagnetic field and the Hawking radiation, stemming from the gravitational horizon.

The presence of two particles in a state, defined by a single wave function leads to serious complications and problems in the usual setting of quantum mechanics whose satisfactory solution requires the formalism of quentum field theory.

1. Equation of motion

To preserve the simple, local nature of the equation of motion in time we separate the two particles an approximate manner by the introduction of a two-dimensional Klein-Gordon spinor

wave function,

$$\begin{pmatrix}\chi_+\\ \chi_-\end{pmatrix} = \frac{1}{2}\begin{pmatrix}\phi + \frac{i}{m}\partial_0\phi\\ \phi - \frac{i}{m}\partial_0\phi\end{pmatrix}. \tag{44}$$

A particle (anti-particle) in rest resides completely in χ_1 (χ_2) and a slow moving particle (anti-particle) should have small χ_2 (χ_1) component. The redoubling of the components of the wave functions halves the order of the differential equation and allows us to write Klein-Gordon equation as

$$i\partial_0\begin{pmatrix}\chi_+\\ \chi_-\end{pmatrix} = \frac{1}{2}\begin{pmatrix}i\partial_0\phi + (m - \frac{1}{m}\boldsymbol{\nabla}^2)\phi\\ i\partial_0\phi + (\frac{1}{m}\boldsymbol{\nabla}^2 - m)\phi\end{pmatrix}. \tag{45}$$

The relations $\chi_+ + \chi_- = \phi$, $\chi_+ - \chi_- = \frac{i}{m}\partial_0\phi$ lead to further simplification,

$$i\partial_0\begin{pmatrix}\chi_+\\ \chi_-\end{pmatrix} = \frac{1}{2}\begin{pmatrix}m(\chi_+ - \chi_-) + (m - \frac{1}{m}\boldsymbol{\nabla}^2)(\chi_+ + \chi_-)\\ m(\chi_+ - \chi_-) + (\frac{1}{m}\boldsymbol{\nabla}^2 - m)(\chi_+ + \chi_-)\end{pmatrix}, \tag{46}$$

which can be written as

$$i\partial_0\chi = H\chi \tag{47}$$

with the Hamiltonian

$$H = -\frac{\boldsymbol{\nabla}^2}{2m}(\sigma_3 + i\sigma_2) + m\sigma_3. \tag{48}$$

The last term on the right hand side sets different rest mass energy for the two components and the first term adds the kinetic energy. The spectrum of the Hamiltonian is $E = \pm\omega_{\boldsymbol{p}}$ and the eigenstate can be identified by the particle and anti-particle states.

Hermiticity The operator H is non-Hermitian, $H^\dagger \neq H$ and as a result its eigenvectors, corresponding to particle and anti-particle are not orthogonal and the unstable. Such a mixing of the particle and anti-particle state is non-physical and we should restore Hermiticity. This is achieved by modifying the scalar product by the help of a "metric operator", g, $\langle\psi|\phi\rangle \to \langle\psi|g|\phi\rangle$. This is formally equivalent with the generalization of each vector $|\psi\rangle$ to a "covariant" and a "contravariant" version, the bra and the ket, respectively. The usual Hermitian conjugation which maps the ket $|\psi\rangle$ into the bra $|\psi\rangle \to |\psi\rangle^\dagger = \langle\psi|$ is replaced by the bra generated by the Klein-Gordon conjugation, $|\psi\rangle \to \langle\bar{\psi}| = \langle\psi|g$ yielding the scalar product $\langle\bar{\psi}|\phi\rangle = \langle\psi|g|\phi\rangle$. The equation $\langle\phi|A^\dagger|\psi\rangle = \langle\psi|A|\phi\rangle^*$ defining the matrix elements of the usual Hermitian conjugate $A^\dagger$ of an operator A, is now replaced by $\langle\bar{\phi}|\bar{A}|\psi\rangle = \langle\bar{\psi}|A|\phi\rangle^*$ written as $\langle\phi|g\bar{A}|\psi\rangle = \langle\psi|gA|\phi\rangle^*$. Hence we

have $g\bar{A} = (gA)^\dagger$, yielding the Klein-Gordon conjugate operator, $\bar{A} = g^{-1}A^\dagger g^\dagger$. We shall use $g = \sigma_3$ which leads to the matrix elements

$$\langle\bar{\chi}|A|\chi'\rangle = \int d^3x\bar{\chi}(\boldsymbol{x})A\chi'(\boldsymbol{x}), \tag{49}$$

and the Klein-Gordon conjugate of wave functions and operators, $\bar{\chi} = \chi^\dagger\sigma_3$ and $A \to \sigma_3 A^\dagger\sigma_3 = \bar{A}$, respectively. It is easy to check that the Hamiltonian is now Hermitean, $\bar{H} = H$.

The modification of the Hermitian conjugation and with it the scalar product influences the physical content, the definition of the expectation values. We are forced to make this momentous step to assure the orthogonality and the stability of the particle and anti-particle eigenstates of H. But this comes with a high price: the scalar product becomes indefinite owing to the non-definite eigenvalues of σ_3. In fact, the naive scalar product for the Klein-Gordon wave function,

$$\langle\phi|\phi'\rangle = \int d^3x\phi^*(t,\boldsymbol{x})\phi'(t,\boldsymbol{x}), \tag{50}$$

is now replaced by

$$\langle\bar{\chi}|\chi'\rangle = \int d^3x\bar{\chi}(t,\boldsymbol{x})\chi'(t,\boldsymbol{x}) = \frac{i}{2m}\int d^3x\phi^*(\boldsymbol{x})\overleftrightarrow{\partial_0}\phi'(\boldsymbol{x}), \tag{51}$$

in particularly the norm of a Klein-Gordon spinor χ is given by the non-definite Noether charge,

$$\langle\bar{\chi}|\chi\rangle = \int d^3x j^0(t,\boldsymbol{x}). \tag{52}$$

The result allows us to keep an important feature of the non-relativistic quantum mechanics, namely that the norm of a state is the space integral of the probability distribution.

Lagrangian The equation of motion (47) can be obtained from the Lagrangian

$$L = \frac{i}{2}\bar{\chi}\partial_0\chi - \frac{i}{2}\partial_0\bar{\chi}\chi - \bar{\chi}H\chi. \tag{53}$$

In fact, the Euler-Lagrange equation, corresponding to the variation of $\bar{\chi}$ is (47) and the variation of χ yields

$$i\partial_0\bar{\chi} = -\bar{\chi}H. \tag{54}$$

Due to $\bar{H} = H$ this equation of motion is equivalent with eq. (47).

2. Free particle

Plane wave: Let us look for the plane wave solution with positive and negative energy,

$$\chi(x) = u_{\boldsymbol{p}}e^{-ipx}, \tag{55}$$

and

$$\chi(x) = v_{\boldsymbol{p}} e^{ipx}. \tag{56}$$

The mass-shell condition, $p^2 = m^2$, requires $p^0 = \omega_p$ and the vectors u and v satisfy the equation

$$\begin{aligned} \omega_p u_{\boldsymbol{p}} &= H_{\boldsymbol{p}} u_{\boldsymbol{p}}, \\ -\omega_p v_{\boldsymbol{p}} &= H_{\boldsymbol{p}} v_{\boldsymbol{p}}, \end{aligned} \tag{57}$$

where the Hamiltonian, acting in a given two dimensional momentum subspace is

$$H_{\boldsymbol{p}} = \frac{\boldsymbol{p}^2}{2m}(\sigma_3 + i\sigma_2) + m\sigma_3. \tag{58}$$

We shall use the eigenvectors

$$\begin{aligned} u_{\boldsymbol{p}} &= \frac{1}{2\sqrt{m\omega_p}} \begin{pmatrix} m + \omega_p \\ m - \omega_p \end{pmatrix} = u_{-\boldsymbol{p}}, \\ v_{\boldsymbol{p}} &= \frac{1}{2\sqrt{m\omega_p}} \begin{pmatrix} m - \omega_p \\ m + \omega_p \end{pmatrix} = v_{-\boldsymbol{p}}, \end{aligned} \tag{59}$$

which are orthogonal,

$$\bar{u}_{\boldsymbol{p}} v_{\boldsymbol{p}} = \bar{v}_{\boldsymbol{p}} u_{\boldsymbol{p}} = 0, \tag{60}$$

and normalized,

$$\bar{u}_{\boldsymbol{p}} u_{\boldsymbol{p}} = -\bar{v}_{\boldsymbol{p}} v_{\boldsymbol{p}} = 1. \tag{61}$$

Wave function: The general solution of the equation of motion (47) can be written as a Fourier integral,

$$\chi(x) = \int_{\boldsymbol{p}} [a_{\boldsymbol{p}} u_{\boldsymbol{p}} e^{-ipx} + b^*_{\boldsymbol{p}} v_{\boldsymbol{p}} e^{ipx}]_{|p^0 = \omega_p}, \tag{62}$$

using the integral measure

$$\int_{\boldsymbol{p}} = \int \frac{d^3 p}{(2\pi)^3}, \tag{63}$$

and the scalar field, corresponding to this solution, is

$$\phi(x) = \int_{\boldsymbol{p}} \sqrt{\frac{m}{\omega_p}} [a_{\boldsymbol{p}} e^{-ipx} + b^*_{\boldsymbol{p}} e^{ipx}]_{|p^0 = \omega_p}. \tag{64}$$

The plane waves of the Fourier decomposition of the general solution are always on mass hell, $p^0 = \omega_p$. The negative energy component, describing the anti-particle content has the three-momentum with the wrong sign. This sign problem can be avoided by interpreting the coefficient function as the complex conjugate of the anti-particle wave function,

$$\begin{aligned} a_{\boldsymbol{p}} &= \frac{i}{2\sqrt{m\omega_{\boldsymbol{p}}}} \int_{\boldsymbol{x}} e^{ipx} \overleftrightarrow{\partial_0} \phi(x), \\ a^*_{\boldsymbol{p}} &= -\frac{i}{2\sqrt{m\omega_{\boldsymbol{p}}}} \int_{\boldsymbol{x}} e^{-ipx} \overleftrightarrow{\partial_0} \phi^*(x), \\ b^*_{\boldsymbol{p}} &= -\frac{i}{2\sqrt{m\omega_{\boldsymbol{p}}}} \int_{\boldsymbol{x}} e^{-ipx} \overleftrightarrow{\partial_0} \phi(x), \\ b_{\boldsymbol{p}} &= \frac{i}{2\sqrt{m\omega_{\boldsymbol{p}}}} \int_{\boldsymbol{x}} e^{ipx} \overleftrightarrow{\partial_0} \phi^*(x). \end{aligned} \tag{65}$$

Note the unusual way both the particle and the anti-particle states are unified within the Klein-Gordon wave function: $\phi(x)$ contains the wave function of the particle and the complex conjugate of the wave function of the anti-particle. The complex conjugation amounts to the flipping of the direction of the time, in agreement with the earlier remark that the time runs in the opposite direction for anti-particles. The wave function, satisfying the second order Klein-Gordon equation may be real, $\phi^*(x) = \phi(x)$, and the anti-particle of a neutral particle is identical with the particle.

Non-relativistic limit: The non-relativistic limit for the particle and the anti-particle components of the scalar field $\phi(x)$ can be found by taking "the square root of the Klein-Gordon equation",

$$i\partial_0 \phi^{(\pm)}(x) = \pm\sqrt{m^2 - \Delta}\,\phi^{(\pm)}(x) \tag{66}$$

yielding

$$i\partial_0 \phi^{(\pm)}(x) = \left[m - \frac{\Delta}{2m} + \mathcal{O}\left(\left(\frac{\Delta}{m} \right)^2 \right) \right] \phi^{(\pm)}(x), \tag{67}$$

and reproduces Schrödinger's equation in the leading order in $\boldsymbol{p}^2/m^2$. The non-relativistic limit for the Klein-Gordon spinor cam be found by noting that χ_2 is suppressed for slow moving particle. This suggests the strategy to eliminate χ_2 by its equation of motion which reduces to an algebraic calcuation in momentum space. We assume a plane wave, $\chi(x) = e^{-ipx}\eta_p$, satisfying

$$p^0 \begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix} = \frac{1}{2} \begin{pmatrix} m(\eta_1 - \eta_2) + (m + \frac{1}{m}\boldsymbol{p}^2)(\eta_1 + \eta_2) \\ m(\eta_1 - \eta_2) - (m + \frac{1}{m}\boldsymbol{p}^2)(\eta_1 + \eta_2) \end{pmatrix}, \tag{68}$$

and use its second component to eliminate η_2,

$$\eta_2 = -\frac{\frac{1}{m}\boldsymbol{p}^2}{2p^0 + 2m + \frac{1}{m}\boldsymbol{p}^2}\eta_1. \tag{69}$$

the resulting effective equation of motion for χ_1,

$$\begin{aligned} 2p^0\eta_1 &= m(\eta_1-\eta_2)+(m+\frac{1}{m}\boldsymbol{p}^2)(\eta_1+\eta_2) \\ &= \left(2m+\frac{1}{m}\boldsymbol{p}^2-\frac{(\frac{1}{m}\boldsymbol{p}^2)^2}{2p^0+2m+\frac{1}{m}\boldsymbol{p}^2}\right)\eta_1, \end{aligned} \tag{70}$$

yields the spectrum, $p^0=m+\epsilon$,

$$\epsilon=\frac{\boldsymbol{p}^2}{2m}-\frac{(\frac{1}{m}\boldsymbol{p}^2)^2}{8m+4\epsilon+\frac{2\boldsymbol{p}^2}{m}}\approx\frac{\boldsymbol{p}^2}{2m}-\frac{(\boldsymbol{p}^2)^2}{8m^2}. \tag{71}$$

However the equations of motion (66) and (70) are not very useful because they are highly non-local in three space.

Projector: The simplest separation of the particle or anti-particle modes is achieved by using the projection operator,

$$\Lambda_{\pm,\boldsymbol{p}}=\begin{cases} u_{\boldsymbol{p}}\otimes\bar{u}_{\boldsymbol{p}} \\ -v_{\boldsymbol{p}}\otimes\bar{v}_{\boldsymbol{p}} \end{cases} =\pm\frac{1}{4m\omega_{\boldsymbol{p}}}\begin{pmatrix}(m\pm\omega_{\boldsymbol{p}})^2 & \boldsymbol{p}^2\\ -\boldsymbol{p}^2 & -(m\mp\omega_{\boldsymbol{p}})^2\end{pmatrix}. \tag{72}$$

The projection operator in coordinate space,

$$\Lambda_\pm(\boldsymbol{x},\boldsymbol{y})=\int_{\boldsymbol{p}}e^{i\boldsymbol{p}(\boldsymbol{x}-\boldsymbol{y})-|\boldsymbol{p}|\epsilon}\Lambda_{\pm,\boldsymbol{p}}, \tag{73}$$

suffers of a singularity as $\boldsymbol{x}-\boldsymbol{y}\to 0$. The fast oscillating contribution of the Fourier integral is suppressed in any application of the projector operator on states with finite length scale according to the Riemann-Lebesbgue lemma, to be taken into account by the introduction of a regulator, the infinitesimal parameter ϵ. The asymptotic behavior

$$\Lambda_{\pm,\boldsymbol{p}}=\begin{cases}\pm\frac{|\boldsymbol{p}|}{4m}(\sigma_3+i\sigma_2) & |\boldsymbol{p}|\gg m,\\ \frac{1}{2}(\mathbb{1}\pm\sigma_3) & |\boldsymbol{p}|\ll m,\end{cases} \tag{74}$$

shows that the low momentum, $|\boldsymbol{p}|\ll m$, particle and the anti-particle amplitude resides mainly in the upper and the lower component of the Klein-Gordon spinor, respectively. The $\boldsymbol{p}$-dependence, displayed at large momenta, $|\boldsymbol{p}|\gg m$, witnesses the mixing of the particle and anti-particle modes in the upper and the lower components.

Charge conjugation: The exchange of particle and anti-particle is performed by the matrix

$$C_{\boldsymbol{p}}=v_{\boldsymbol{p}}\otimes\bar{u}_{\boldsymbol{p}}-u_{\boldsymbol{p}}\otimes\bar{v}_{\boldsymbol{p}}=\frac{1}{4m\omega_{\boldsymbol{p}}}\left[\begin{pmatrix}m^2-\omega_{\boldsymbol{p}}^2 & -(m-\omega_{\boldsymbol{p}})^2\\ (m+\omega_{\boldsymbol{p}})^2 & \omega_{\boldsymbol{p}}^2-m^2\end{pmatrix}-\begin{pmatrix}m^2-\omega_{\boldsymbol{p}}^2 & -(m+\omega_{\boldsymbol{p}})^2\\ (m-\omega_{\boldsymbol{p}})^2 & \omega_{\boldsymbol{p}}^2-m^2\end{pmatrix}\right]=\sigma_1, \tag{75}$$

in the Klein-Gordon spinor space within a given momentum subspace.

Indefinite norm: The norm of the state, defined by the wave-function (62)

$$\begin{aligned}\langle\chi|\chi\rangle &= \int d^3x\bar{\chi}(x)\chi(x)\\ &= \int_{\boldsymbol{p}}(a^*_{\boldsymbol{p}}a_{\boldsymbol{p}} - b_{\boldsymbol{p}}b^*_{\boldsymbol{p}}),\end{aligned} \tag{76}$$

is non-definite, anti-particles have negative norm and are in conflict with the probabilistic interpretation of the wave function. One could have guessed this problem, the non-definite nature of the probability density, at the very beginning by noting that a conserved current, $\partial_\mu j^\mu = 0$ must be linear in the first derivative of the scalar wave function $\phi(x)$ and the density, j^0, being proportional to $\partial_0\phi$ and $\partial\phi^*$ have different sign for particle and anti-particle wave functions.

Expectation values: The definition of the expectation value is left open in a linear space with indefinite norm, we may use either

$$\langle A\rangle = \frac{\langle\chi|A|\chi\rangle}{|\langle\chi|\chi\rangle|}, \tag{77}$$

or

$$\langle A\rangle = \frac{\langle\chi|A|\chi\rangle}{\langle\chi|\chi\rangle}. \tag{78}$$

The expectation value of the energy-momentum operator, $p^\mu = (H, -i\boldsymbol{\nabla})$, is

$$\begin{aligned}\langle\chi|p^\mu|\chi\rangle &= \int d^3x\bar{\chi}(t,\boldsymbol{x})(H,-i\boldsymbol{\nabla})\chi(t,\boldsymbol{x})\\ &= \int d^3x\int_{\boldsymbol{pq}}[a^*_{\boldsymbol{p}}\bar{u}_{\boldsymbol{p}}e^{ipx} + b_{\boldsymbol{p}}\bar{v}_{\boldsymbol{p}}e^{-ipx}](\omega_q,\boldsymbol{q})[a_{\boldsymbol{q}}u_{\boldsymbol{q}}e^{-iqx} - b^*_{\boldsymbol{q}}v_{\boldsymbol{q}}e^{iqx}]\\ &= \int_{\boldsymbol{p}}(\omega_p,\boldsymbol{p})[a^*_{\boldsymbol{p}}a_{\boldsymbol{p}} + b_{\boldsymbol{p}}b^*_{\boldsymbol{p}}].\end{aligned} \tag{79}$$

The Noether current, (35), with $A_\mu = 0$, in a state, described by the wave function $\phi(x)$ is the expectation value of the operator $i\overleftrightarrow{\partial}^\mu/2m$. The relations

$$\begin{aligned}j_0 &= \frac{1}{2}(\chi_+ + \chi_-)^*(\chi_+ - \chi_-) + \frac{1}{2}(\chi_+ - \chi_-)^*(\chi_+ + \chi_-) = \bar{\chi}(x)\chi(x),\\ \boldsymbol{j} &= -\frac{i}{2m}(\chi_+ + \chi_-)^*\overleftrightarrow{\boldsymbol{\nabla}}(\chi_+ + \chi_-) = -\frac{i}{2m}\bar{\chi}(x)\overleftrightarrow{\boldsymbol{\nabla}}(\sigma_3 + i\sigma_2)\chi(x),\end{aligned} \tag{80}$$

the minus sign in the last equation is due to the relation $j_\mu = (j^0, -\boldsymbol{j})$, show that $\langle\chi|j^0|\chi\rangle$ is given by the norm, (76), and

$$\begin{aligned}\langle\chi|\boldsymbol{j}|\chi\rangle &= -\frac{i}{2m}\int d^3x\int_{\boldsymbol{pq}}[a^*_{\boldsymbol{p}}\bar{u}_{\boldsymbol{p}}e^{ipx} + b_{\boldsymbol{p}}\bar{v}_{\boldsymbol{p}}e^{-ipx}]\overleftrightarrow{\boldsymbol{\nabla}}(\sigma_3 + i\sigma_2)[a_{\boldsymbol{q}}u_{\boldsymbol{q}}e^{-iqx} + b^*_{\boldsymbol{q}}v_{\boldsymbol{q}}e^{iqx}]\\ &= \Re\frac{1}{m}\int d^3x\int_{\boldsymbol{pq}}\boldsymbol{q}[a^*_{\boldsymbol{p}}\bar{u}_{\boldsymbol{p}}e^{ipx} + b_{\boldsymbol{p}}\bar{v}_{\boldsymbol{p}}e^{-ipx}](\sigma_3 + i\sigma_2)[a_{\boldsymbol{q}}u_{\boldsymbol{q}}e^{-iqx} - b^*_{\boldsymbol{q}}v_{\boldsymbol{q}}e^{iqx}]\end{aligned}$$

$$= \int_{\boldsymbol{p}} \frac{\boldsymbol{p}}{m}[a^*_{\boldsymbol{p}} a_{\boldsymbol{p}} \bar{u}_{\boldsymbol{p}}(\sigma_3 + i\sigma_2)u_{\boldsymbol{p}} - b_{\boldsymbol{p}} b^*_{\boldsymbol{p}} \bar{v}_{\boldsymbol{p}}(\sigma_3 + i\sigma_2)v_{\boldsymbol{p}}]. \tag{81}$$

The straightforward calculation,

$$\begin{aligned}
\bar{u}_{\boldsymbol{p}}(\sigma_3 + i\sigma_2)u_{\boldsymbol{p}} &= \frac{1}{4m\omega_{\boldsymbol{p}}}(m + \omega_{\boldsymbol{p}}, -m + \omega_{\boldsymbol{p}}) \begin{pmatrix} 1 & 1 \\ -1 & -1 \end{pmatrix} \begin{pmatrix} m + \omega_{\boldsymbol{p}} \\ m - \omega_{\boldsymbol{p}} \end{pmatrix} \\
&= \frac{(m + \omega_{\boldsymbol{p}}, -m + \omega_{\boldsymbol{p}})}{2\omega_{\boldsymbol{p}}} \begin{pmatrix} 1 \\ -1 \end{pmatrix} \\
&= \frac{m}{\omega_{\boldsymbol{p}}} \\
&= \bar{v}_{\boldsymbol{p}}(\sigma_3 + i\sigma_2)v_{\boldsymbol{p}},
\end{aligned} \tag{82}$$

yields

$$\langle\chi|j^\mu|\chi\rangle = \int_{\boldsymbol{p}} \left(1, \frac{\boldsymbol{p}}{\omega_p}\right)[a^*_{\boldsymbol{p}} a_{\boldsymbol{p}} - b_{\boldsymbol{p}} b^*_{\boldsymbol{p}}]. \tag{83}$$

Note that whatever normalization is chosen for the expectation values, either the energy is unbounded from below in eq. (79) or the probability density is non-definite in eq. (83). The choice (77) stabilizes the energy and leads to ground state. The non-definite nature of j^0 can be understood by recalling that the Noether current is weighted by the electric charge which changes signs when particles and anti-particles are exchanged. One can interpret the dynamics as long as we have exclusively particles or anti-particles. Such a restriction applies anyhow to a description based on a wave function with fixed number of variables. In fact, when particles and anti-particles are present simultaneously then annihilation may take place which requires the reduction of the number of degrees of freedom and correspondingly the decrease of the number of the variables of the wave function. But the separation of the particle and the anti-particle modes can not be maintained in the presence of external electromagnetic field as we shall see below.

3. Localization

There are simple arguments, suggesting problems with localizing particles or anti-particles:

1. **Spread of the wave packet:** Consider a non-relativistic wave packet of width Δx. The speed v_{spr} of the spread of a wave packet can be estimated by the help of the Heisenberg uncertainty relation, $v_{spr} \sim \Delta p/m \sim 1/m\Delta x$. One expects that velocity of the spread will be bounded by the speed of light in relativistic quantum mechanics, requiring a non-local realization of the coordinate operator.

2. **Pair creation:** It is well known that if a one-dimensional particle is confined into an interval of length Δx then it develops a discrete momentum spectrum, $p_n = 2\pi n/\Delta x$. As the localization becomes strong, $\Delta x \to 0$, a decay to a lower stationary state, $n \to n-1$, provides the energy, $\Delta E = \sqrt{m^2 + p_{n+1}^2} - \sqrt{m^2 + p_n^2}$, sufficient to create a particle-anti particle pair. This process starts at the localization which is comparable with the Compton wavelength, $\Delta x = \lambda_c = 1/m$. In other words, a particle may emit particle-anti particle radiation when localized in a smaller region than its Compton wavelength.

Both arguments indicate that there is a maximal localization of a relativistic particle at around the size of its Compton wavelength.

In fact, the separation of the particle and the anti-particle mode was partially carried out by the introduction of the two component wave function (44). The full separation which requires infinitely long observation time can nevertheless be carried out locally in time for free particles, cf. the spinors (59) for $\boldsymbol{p} \neq 0$, by the help of the equation of motion which relates the space- and time-dependence. The particle and anti-particle states of a free particle is defined for each momentum by eqs. (59) and the matrix

$$S_{\boldsymbol{p}} = \frac{m + \omega_p - \sigma_1(m - \omega_p)}{2\sqrt{m\omega_p}} \tag{84}$$

brings the original, Klein-Gordon basis into the new, momentum-independent one,

$$\begin{aligned} S_{\boldsymbol{p}} u_{\boldsymbol{p}} &= \frac{mc + \omega_p - \sigma_1(mc - \omega_p)}{2\sqrt{mc\omega_p}} \frac{1}{2\sqrt{mc\omega_p}} \begin{pmatrix} mc + \omega_p \\ mc - \omega_p \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} = w_+ \\ S_{\boldsymbol{p}} v_{\boldsymbol{p}} &= \frac{mc + \omega_p - \sigma_1(mc - \omega_p)}{2\sqrt{mc\omega_p}} \frac{1}{2\sqrt{mc\omega_p}} \begin{pmatrix} mc - \omega_p \\ mc + \omega_p \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix} = w_-. \end{aligned} \tag{85}$$

This is a unitary transformation,

$$\bar{S}_{\boldsymbol{p}} = \sigma_3 S^\dagger \sigma_3 = \frac{m + \omega_p + \sigma_1(m - \omega_p)}{2\sqrt{m\omega_p}} = S_{\boldsymbol{p}}^{-1}, \tag{86}$$

allows us to write the Hamiltonian within the sector $\boldsymbol{p}$ in the form

$$H_{\boldsymbol{p}} = S_{\boldsymbol{p}}^{-1}[w_+ \omega_{\boldsymbol{p}} \bar{w}_+ - w_-(-\omega_{\boldsymbol{p}})\bar{w}_-] S_{\boldsymbol{p}} \tag{87}$$

where

$$H_{FV} = S_{\boldsymbol{p}} H_{\boldsymbol{p}} S_{\boldsymbol{p}}^{-1} = \omega_{\boldsymbol{p}}(w_+ \bar{w}_+ + w_- \bar{w}_-) = \sigma_3 \omega_p, \tag{88}$$

is a diagonal matrix. The explicit calculation of the similarity transformation is the simplest by using $\{\sigma_a, \sigma_b\} = 2\delta_{a,b}$, c.f. eq. (163),

$$\begin{aligned} H_{FV} &= S_{\boldsymbol{p}} H_{\boldsymbol{p}} S_{\boldsymbol{p}}^{-1} \\ &= H_{\boldsymbol{p}} \left(\frac{m + \omega_p + \sigma_1(m - \omega_p)}{2\sqrt{m\omega_p}} \right)^2 \\ &= \sigma_3 \omega_p. \end{aligned} \tag{89}$$

The Hamiltonian obviously possesses the right non-relativistic limit.

The momentum-dependence of the basis transformation, $S_{\boldsymbol{p}}$, can easily be understood in the following manner. The change of the momentum $\boldsymbol{p}$ induces a change of ω_p and the Klein-Gordon spinor $\chi_\pm(x) = (1 \pm \frac{p^0}{m}) e^{-ipx}$, requiring a modified basis transformation S.

The transformation $S_{\boldsymbol{p}}$ can be extended from a given momentum sector to the whole Hilbert space,

$$\begin{aligned} \langle \boldsymbol{p} | S_{aa'} | \boldsymbol{p}' \rangle &= (2\pi)^3 \delta(\boldsymbol{p} - \boldsymbol{p}') S_{\boldsymbol{p}aa'}, \\ \langle \boldsymbol{x} | S_{aa'} | \boldsymbol{x}' \rangle &= \int_{\boldsymbol{p}\boldsymbol{p}'} \langle \boldsymbol{x} | \boldsymbol{p} \rangle \langle \boldsymbol{p} | S_{aa'} | \boldsymbol{p}' \rangle \langle \boldsymbol{p}' | \boldsymbol{x}' \rangle = \int_{\boldsymbol{p}} S_{\boldsymbol{p}aa'} e^{i\boldsymbol{p}(\boldsymbol{x} - \boldsymbol{x}')}, \end{aligned} \tag{90}$$

taking the general solution of the Klein-Gordon equation, (62), into

$$\chi_{FV}(x_{FV}) = S\chi_{KG}(x_{KG}) = \int_{\boldsymbol{p}} [a_{\boldsymbol{p}} e^{-ipx_{FV}} w_+ + b^*_{\boldsymbol{p}} e^{ipx_{FV}} w_-] = \begin{pmatrix} \chi_+(x_{FV}) \\ \chi_-(x_{FV}) \end{pmatrix} \tag{91}$$

in the decoupled, Feshbach-Villars basis where $\chi_+(x_{FV})$ and $\chi_-(x_{FV})$ stand for the particle and anti-particle wave function.

It is a point of central importance that a momentum-dependent mixing of the Klein-Gordon spinor components, described by the matrix S, obtained by extending $S_{\boldsymbol{p}}$ to all momentum subspace, changes the definition of the coordinate operator, $\boldsymbol{x}_{KG} \to \boldsymbol{x}_{FV}$, indicated explicitely in (91). In fact, the representation $\hat{\boldsymbol{x}}_{KG} = i\boldsymbol{\nabla}_{p_{KG}}$ of the coordinate operator yields

$$\hat{\boldsymbol{x}}_{FV} = S\hat{\boldsymbol{x}}_{KG}\bar{S} = Si\boldsymbol{\nabla}_p\bar{S} \tag{92}$$

and the application of the identity, $[i\boldsymbol{\nabla}_p, f_{\boldsymbol{p}}] = i\boldsymbol{\nabla}_p f_{\boldsymbol{p}}$, gives

$$\hat{\boldsymbol{x}}_{FV} = i(\boldsymbol{\nabla}_p + \boldsymbol{c}_{\boldsymbol{p}}\sigma_1) = \hat{\boldsymbol{x}}_{KG} + i\boldsymbol{c}_{\boldsymbol{p}}\sigma_1, \tag{93}$$

with $\boldsymbol{c}_{\boldsymbol{p}} = -\boldsymbol{\nabla}_p\omega_p/2\omega_p$. The first term on the right hand side is the usual non-relativistic expression and the second, "connection" term results from the $\boldsymbol{p}$-dependence of the eigenvectors $u_{\boldsymbol{p}}$, $v_{\boldsymbol{p}}$ and mixes the positive and the negative frequency components. This construction is formally similar

to an Abelian gauge field, introduced in momentum rather than coordinate space. The important lesson is that the coordinate operator of the original, Klein-Gordon representation mixes the (free) particle and anti-article states, opening up the possibility of creation and annihilation of particle-anti particle pairs by the electromagnetic field. This is not an entirely unexpected effect since a measue of the coordinate is an attempt to localise the particle beyond its Compton wavelength.

This problem, namely that the physical role of the variables of $\chi_{FV}(x)$ and $\chi_{FV}(x)$ are different in eq. (91), is hidden by the notation which assigns the same symbol to the variable of both functions. In fact, the components $\boldsymbol{x}$ denotes a spectrum element of the coordinate operator and the spectrum of $\hat{\boldsymbol{x}}_{KG}$ and $\hat{\boldsymbol{x}}_{VS}$ are identical. To make the notation clearer we introduce the coordinate and momentum eigenstates,

$$\hat{\boldsymbol{x}}_{KG}|\boldsymbol{x},a\rangle = \boldsymbol{x}|\boldsymbol{x},a\rangle, \quad \hat{\boldsymbol{p}}_{KG}|\boldsymbol{p},a\rangle = \boldsymbol{p}|\boldsymbol{p},a\rangle, \quad \hat{\boldsymbol{x}}_{FV}|\dot{\boldsymbol{x}},\dot{a}\rangle = \dot{\boldsymbol{x}}|\dot{\boldsymbol{x}},\dot{a}\rangle, \quad \hat{\boldsymbol{p}}_{FV}|\dot{\boldsymbol{p}},\dot{a}\rangle = \dot{\boldsymbol{p}}|\dot{\boldsymbol{p}},\dot{a}\rangle, \tag{94}$$

where the quantum numbers with a dot belong to the decoupled basis. The overlap between the coordinate and the momentum bases is given by the usual expression,

$$\langle \boldsymbol{x},a|\boldsymbol{p},b\rangle = a\delta_{ab}e^{i\boldsymbol{x}\boldsymbol{p}}, \quad \langle \dot{\boldsymbol{x}},\dot{a}|\dot{\boldsymbol{p}},\dot{b}\rangle = \dot{a}\delta_{\dot{a}\dot{b}}e^{i\dot{\boldsymbol{x}}\dot{\boldsymbol{p}}}, \tag{95}$$

however transfer between the coupled and the decoupled bases is given by

$$\langle \dot{\boldsymbol{p}},\dot{a}|\boldsymbol{p},a\rangle = (2\pi)^3\delta(\dot{\boldsymbol{p}}-\boldsymbol{p})S_{\boldsymbol{p}\dot{a}a}, \quad \langle \boldsymbol{p},a|\dot{\boldsymbol{p}},\dot{a}\rangle = (2\pi)^3\delta(\dot{\boldsymbol{p}}-\boldsymbol{p})\bar{S}_{\boldsymbol{p}a\dot{a}}. \tag{96}$$

The general solution (91) reads as

$$\chi(\dot{x}) = \int_{\dot{\boldsymbol{p}}}[a_{\dot{\boldsymbol{p}}}w_+e^{-i\dot{p}\dot{x}} + b^*_{\dot{\boldsymbol{p}}}w_-e^{i\dot{p}\dot{x}}]. \tag{97}$$

The limitation on the localizability of a particle or anti-particle is an inextricable difficulty in recovering the usual formalism in the quantum mechanics for relativistic particles and forces us to make a radical step, to change the position operator which up to now has been taken over naively from the non-relativistic case. We shall check that the spread of the wave-packet can consistently be interpreted in the decoupled basis.

The completeness relations,

$$\mathbb{1} = \sum_a \int d^3x|\boldsymbol{x},a\rangle a\langle \boldsymbol{x},a| = \sum_a \int_{\boldsymbol{p}} |\boldsymbol{p},a\rangle a\langle \boldsymbol{p},a| = \sum_{\dot{a}} \int_{\dot{\boldsymbol{p}}} |\dot{\boldsymbol{p}},\dot{a}\rangle\dot{a}\langle \dot{\boldsymbol{p}},\dot{a}|, \tag{98}$$

allows us to write the decoupled coordinate eigenstate in the original basis,

$$|\dot{\boldsymbol{x}},\dot{a}\rangle \;=\; \sum_a \int_{\boldsymbol{p}\dot{\boldsymbol{p}}} d^3x|\boldsymbol{x},a\rangle a\langle \boldsymbol{x},a|\boldsymbol{p},a\rangle a\langle \boldsymbol{p},a|\dot{\boldsymbol{p}},\dot{a}\rangle\dot{a}\langle \dot{\boldsymbol{p}},\dot{a}|\dot{\boldsymbol{x}},\dot{a}\rangle$$

$$= \sum_a \int_{\boldsymbol{p}} d^3x |\boldsymbol{x}, a\rangle \bar{S}_{\boldsymbol{p}a\dot{a}} \dot{a} e^{i(\boldsymbol{x}-\dot{\boldsymbol{x}})\boldsymbol{p}}, \tag{99}$$

leading to the overlap

$$\langle \dot{\boldsymbol{x}}, \dot{a} | \boldsymbol{x}, a\rangle = \int_{\boldsymbol{p}} \dot{a} S_{\boldsymbol{p}\dot{a}a} e^{i(\dot{\boldsymbol{x}}-\boldsymbol{x})\boldsymbol{p}}, \quad \langle \boldsymbol{x}, a | \dot{\boldsymbol{x}}, \dot{a}\rangle = \int_{\boldsymbol{p}} \bar{S}_{\boldsymbol{p}a\dot{a}} \dot{a} e^{i(\boldsymbol{x}-\dot{\boldsymbol{x}})\boldsymbol{p}}, \tag{100}$$

in the coordinate basis.

4. The birth of relativistic quantum field theory

The two problems, mentioned at the end of section (II A), can be solved by building up another representation of the physical states and observables than in Schrödinger's quantum mechanics and its relativistic extension, discussed so far. The starting point is the observation that the spectrum of a system of particles with a given momentum, $\dot{\boldsymbol{p}}$, is equidistant, $E_n = n\omega_{\dot{\boldsymbol{p}}}$, n denoting the number of particles. The only one dimensional quantum system with such a spectrum is the harmonic oscillator with frequency $\omega_{\dot{\boldsymbol{p}}}$. Thus one defines a harmonic oscillator for the particles and anti-particles within each momentum sector by introducing the formal variables $X_{\dot{\boldsymbol{p}},\dot{a}}$, $P_{\dot{\boldsymbol{p}},\dot{a}}$ with the usual commutation relation, $[X_{\dot{\boldsymbol{p}},\dot{a}}, P_{\dot{\boldsymbol{p}}',\dot{a}'}] = i\delta_{\dot{a},\dot{a}'}\delta(\dot{\boldsymbol{p}} - \dot{\boldsymbol{p}}')$ and Hamiltonian is additive for noninteracting particles,

$$H = \sum_{\dot{a}} \int_{\dot{\boldsymbol{p}}} \left(\frac{1}{2} P^2_{\dot{\boldsymbol{p}},\dot{a}} + \frac{\omega^2_{\dot{\boldsymbol{p}}}}{2} X^2_{\dot{\boldsymbol{p}},\dot{a}} \right). \tag{101}$$

One can imagine such a system of infinitely many harmonic oscillator as a series of boxes, assigned to each possible value of the three-momentum. The boxes may contain balls, representing the particles, the box corresponding to $\dot{\boldsymbol{p}}$ including $n_{\dot{\boldsymbol{p}}}$ balls. We actually need two boxes for each three-momentum, one for the particles and the other for the anti-particles. The multi-particle states are given by the help of such a double occupation number, $|n_{\dot{\boldsymbol{p}}}, \bar{n}_{\dot{\boldsymbol{p}}}\rangle$ where $n_{\dot{\boldsymbol{p}}}$ and $\bar{n}_{\dot{\boldsymbol{p}}}$ denotes the number of particles and anti-particles, respectively, in the bos of momentum $\dot{\boldsymbol{p}}$.

The quantum field, an x-dependent operator, is obtained from (97) where we dispose the spinors $w_\pm$ by using the operators

$$a_{\dot{\boldsymbol{p}}} = \frac{\omega_{\dot{\boldsymbol{p}}} X_{\dot{\boldsymbol{p}},+} + iP_{\dot{\boldsymbol{p}},+}}{\sqrt{2\omega_{\dot{\boldsymbol{p}}}}}, \quad b_{\dot{\boldsymbol{p}}} = \frac{\omega_{\dot{\boldsymbol{p}}} X_{\dot{\boldsymbol{p}},-} + iP_{\dot{\boldsymbol{p}},-}}{\sqrt{2\omega_{\dot{\boldsymbol{p}}}}}. \tag{102}$$

The resulting expression,

$$\chi(\dot{x}) = \int_{\dot{\boldsymbol{p}}} [a_{\dot{\boldsymbol{p}}} e^{-i\dot{p}\dot{x}} + b^\dagger_{\dot{\boldsymbol{p}}} e^{i\dot{p}\dot{x}}], \tag{103}$$

contains operator valued Fourier coefficients. One could have use any other combination of $X_{\dot{\boldsymbol{p}},\dot{a}}$ and $P_{\dot{\boldsymbol{p}},\dot{a}}$ in constructing a quantum field but it is known that the creation and annihilation operators offers the simplest and clearest equations for quantum harmonic oscillators. The quantum field (103) removes a particle and creates and anti-particle and the spinors $w_\pm$ are left out because the operators $a_{\boldsymbol{p}}$ and $b_{\boldsymbol{p}}$ act on Hilbert spaces associated with w_+ and w_-, respectively.

Note that the field (103) looks superficially than as a wave function, the general solution of the free equation of motion except that it is operator valued function of the space-time coordinates. This is the origin of the name "second quantization": The first quantization, the introduction of the wave function in Schrödinger's formalism is followed by the replacement of the c-number valued wave functin by operators, a second quantization procedure. The key differences between the two quantization procedure, the different origin of the Hilbert space, the scalar product and the observables, in particular the Hamiltonian, make it possible to avoid the problems, mentioned at the end of section (II A). The Hamiltonian (101) has obviously positive spectrum and the problem of the negative anti-particle energy is cured. the price is to rely on the energy, (101), defined by the harmonic oscillators rather than the time dependence in the quantum field. The probability distribution is positive definite for harmonic oscillators hence the problem with the non-definite nature of the Noether current is eliminated, too. Both results originate algebraically from the non-commutativity of the Fourier coefficients, (102),

$$[a_{\dot{\boldsymbol{p}}}, a^\dagger_{\dot{\boldsymbol{p}}'}] = [b_{\dot{\boldsymbol{p}}}, b^\dagger_{\dot{\boldsymbol{p}}'}] = (2\pi)^3\delta(\dot{\boldsymbol{p}} - \dot{\boldsymbol{p}}'). \tag{104}$$

The representation of the multi-particle states with the occupation number, $|n_{\dot{\boldsymbol{p}}}, \bar{n}_{\dot{\boldsymbol{p}}}\rangle$, solves yet another, technical problem. Namely, the multi-particle states must be symmetrized with respect to the exchange of particles. This forces us to use multi-particle states where the symmetry is achieved by summing over the permutations of the variables of the wave function. Such a sum contains $n!$ contribution for an n-particle state and renders the use of such a state extremely difficult beyond few particles. The particles with the same momentum are represented by the different quanta of excitations of a given harmonic oscillator. The n excitations enter in an indistinguishable manner in the state $a^{\dagger n}|0\rangle$ and the symmetrization is automatically achieved.

The former similarity of (97) and (103) suggest the interpretation of the quantum field as some kind of generalization of the one-particle wave function. The generalization involves the replacement of c-numbers, the Fourier amplitudes, $a_{\dot{\boldsymbol{p}}}$ and $b^*_{\dot{\boldsymbol{p}}}$, by operators, reminiscent of the quantization of a classical system. Since this manipulation is performed on a wave function, resulting from a quantization procedure, the appearance of the quantum field was historically

associated with second quantization, by changing the continuous spectrum of a harmonic oscillator to discrete one. This is a misleading analogy because it holds for free particles only, the reason being that lack of wave function, associated to a part of an interacting system. In fact, the system-environment entanglement generates mixed system states and requires the use of density matrix rather than state vectors. The picture of second quantization, taken more seriously, requires the use of bi-local fields, corresponding to one-particle density matrices.

The difference between the first and the second quantized formalism can better be seen by considering the wave function as a map of the space-time, called external space, into C^n where n is the number of the components of the wave functions ($n = 2$ for χ), called internal space. The name can be justified by regarding the space-time and the value of the wave function as an external or internal structure from the point of view of an elementary particle. The original quantum mechanics is based on quantization rules in the external space, e.g. the replacement of the classical dynamical variables, the functions of the external space coordinates and their canonical pairs, by operators. The second quantization keeps the external variable as (quantum) numbers and applies the quantization rules within the internal space and replaces the value of the wave function and its canonical pair with operators. The Hilbert space of physical states and the scalar product are constructed in a different manner than in the first quantized theory.

The by now standard notations in quantum field theory is simplified and the dots are left out from the equations, written in the decoupled basis. This step contains the danger of confusion and one always has to remember that all equations of relativistic quantum field theory are given in the decoupled basis where the coordinate operator is different than that in the non-relativistic formalism. This is all the more subtle point because a spinless particle is described by the scalar, Klein-Gordon field, $\phi(x)$, without referring to the first order formalism however the separation of the particle and the anti-particle modes, assumed tacitly by using different harmonic oscillators for them, relies on the decoupled basis.

The development of these ideas and the systematical build up of relativistic quantum field theories go beyond this lecture and we shall restrict our attention to some remarks about it and continue with the presentation of the first quantized, relativistic quantum mechanics with fixed number of degrees of freedom.

5. Spread of the wave packet

The problems with localization can the clearest be seen by considering the spread of a wave-packet. A wave function of a wave-packet of a non-relativistic particle is of the form,

$$\psi(t,\boldsymbol{x}) = \int_{\boldsymbol{p}} \psi_{\boldsymbol{p}} e^{-itE_p + i\boldsymbol{x}\boldsymbol{p}}, \tag{105}$$

where $E_{\boldsymbol{p}} = \boldsymbol{p}^2/2m$. The expectation value of the coordinate is

$$\begin{aligned}
\langle \boldsymbol{x} \rangle &= \int_{\boldsymbol{p}\boldsymbol{q}} d^3x \boldsymbol{x} \psi^*_{\boldsymbol{p}} \psi_{\boldsymbol{q}} e^{-it(E_q - E_p) + i\boldsymbol{x}(\boldsymbol{q}-\boldsymbol{p})} \\
&= \int_{\boldsymbol{p}\boldsymbol{q}} d^3x \psi^*_{\boldsymbol{p}} \psi_{\boldsymbol{q}} e^{it(E_p - E_q)} (-i\boldsymbol{\nabla}_q) e^{i\boldsymbol{x}(\boldsymbol{q}-\boldsymbol{p})} \\
&= \int_{\boldsymbol{p}} \psi^*_{\boldsymbol{p}} (i\boldsymbol{\nabla}_p + t\boldsymbol{\nabla} E_p) \psi_{\boldsymbol{p}} \\
&= \langle \psi | \boldsymbol{x} | \psi \rangle_0 + t \langle \psi | \boldsymbol{v}_{gr} | \psi \rangle_0,
\end{aligned} \tag{106}$$

where a partial integration was carried out in arriving at the third equation and the subscript 0 indicates that the matrix elements are calculated between the states which are taken at $t = 0$. One finds a free particle trajectory on the level of the expectation value, starting at the expectation value of the coordinate operator, $\hat{\boldsymbol{x}} = i\boldsymbol{\nabla}_p$, at the initial time, in agreement with Ehrenfest's theorem. The velocity, expectation value of the group velocity, $\boldsymbol{v}_{gr}(\boldsymbol{p}) = \boldsymbol{\nabla} E_p = \boldsymbol{p}/m$, is extracted by the coordinate operator acting on the time dependence of the plane waves.

To follow the spread of the wave packet we need the expectation value of the square of the coordinate,

$$\begin{aligned}
\langle \boldsymbol{x}^2 \rangle &= -\int_{\boldsymbol{p}\boldsymbol{q}} d^3x \psi^*_{\boldsymbol{p}} \psi_{\boldsymbol{q}} e^{it(E_p - E_q)} \boldsymbol{\nabla}^2_q e^{-i\boldsymbol{x}(\boldsymbol{p}-\boldsymbol{q})} \\
&= -\int_{\boldsymbol{p}} \left[\psi^*_{\boldsymbol{p}} e^{itE_p} \boldsymbol{\nabla}^2 \left(\psi_{\boldsymbol{p}} e^{-itE_p} \right) \right] \\
&= -\int_{\boldsymbol{p}} \psi^*_{\boldsymbol{p}} [\boldsymbol{\nabla}^2 - 2it\boldsymbol{\nabla} E_p \boldsymbol{\nabla} - it\boldsymbol{\nabla}^2 E_p - t^2 (\boldsymbol{\nabla} E_p)^2] \psi_{\boldsymbol{p}} \\
&= \langle \psi | (\boldsymbol{x} + t\boldsymbol{v}_{gr})^2 | \psi \rangle_0.
\end{aligned} \tag{107}$$

The spread is the second moment of the coordinate, $\sigma^2_x(t) = \langle x^2 \rangle - \langle x \rangle^2$, is given by

$$\begin{aligned}
\sigma^2_x(t) &= \langle \psi | (\boldsymbol{x} + t\boldsymbol{v}_{gr})^2 | \psi \rangle_0 - (\langle \psi | \boldsymbol{x} | \psi \rangle_0 + t \langle \psi | \boldsymbol{v}_{gr} | \psi \rangle_0)^2 \\
&= \sigma^2_x(0) + t(\langle \psi | \boldsymbol{x}\boldsymbol{v}_{gr} + \boldsymbol{v}_{gr}\boldsymbol{x} | \psi \rangle_0 - 2 \langle \psi | \boldsymbol{x} | \psi \rangle_0 \langle \psi | \boldsymbol{v}_{gr} | \psi \rangle_0) + t^2 \sigma^2_v(0)
\end{aligned} \tag{108}$$

where

$$\sigma^2_v(0) = \langle \psi | \boldsymbol{v}^2_{gr} | \psi \rangle_0 - \langle \psi | \boldsymbol{v}_{gr} | \psi \rangle^2_0 \tag{109}$$

denotes the spread of the velocity. Note that the particle can be localized with arbitrary precision, ie. $\sigma_x^2(0)$ can be arbitrarily small and the speed of the spreading, $\partial_t\sigma_x^2(t)$, can be arbitrarily large because $\boldsymbol{v}_{gr}$ is an unbounded operator.

We turn now to the relativistic case where one expects that velocity of the motion and the speed will be limited and the component of the coordinate operator which mixes particles and anti-particles generates qualitatively new contributions compared to the non-relativistic motion. We carry out the calculation first in the decoupled basis which is formally more similar to the non-relativistic case. The average of the coordinate of the decoupled basis,

$$\begin{aligned}
\langle\chi|\boldsymbol{x}_{FV}|\chi\rangle &= i\int_{\dot{\boldsymbol{p}}\dot{\boldsymbol{q}}} d^3\dot{x}[a^*_{\dot{\boldsymbol{p}}}\bar{w}_+e^{it\omega_{\dot{p}}-i\dot{\boldsymbol{x}}\dot{\boldsymbol{p}}}+b_{\dot{\boldsymbol{p}}}\bar{w}_-e^{-it\omega_{\dot{p}}+i\dot{\boldsymbol{x}}\dot{\boldsymbol{p}}}]\\
&\quad [-a_{\dot{\boldsymbol{q}}}w_+e^{-it\omega_{\dot{q}}}\boldsymbol{\nabla}_{\dot{q}}e^{i\dot{\boldsymbol{x}}\dot{\boldsymbol{q}}}+b^*_{\dot{\boldsymbol{q}}}w_-e^{it\omega_{\dot{p}}}\boldsymbol{\nabla}_{\dot{q}}e^{-i\dot{\boldsymbol{x}}\dot{\boldsymbol{q}}}]\\
&= \int_{\dot{\boldsymbol{q}}}[a^*_{\dot{\boldsymbol{q}}}(i\boldsymbol{\nabla}+t\boldsymbol{\nabla}\omega_{\dot{q}})a_{\dot{\boldsymbol{q}}}+b_{\dot{\boldsymbol{q}}}(i\boldsymbol{\nabla}-t\boldsymbol{\nabla}\omega_{\dot{q}})b^*_{\dot{\boldsymbol{q}}}]\\
&= \int_{\dot{\boldsymbol{q}}}[a^*_{\dot{\boldsymbol{q}}}(i\boldsymbol{\nabla}+t\boldsymbol{\nabla}\omega_{\dot{q}})a_{\dot{\boldsymbol{q}}}-b^*_{\dot{\boldsymbol{q}}}(i\boldsymbol{\nabla}+t\boldsymbol{\nabla}\omega_{\dot{q}})b_{\dot{\boldsymbol{q}}}]\\
&= \langle a|\boldsymbol{x}_{FV}+t\boldsymbol{v}_{gr}|a\rangle_0-\langle b|\boldsymbol{x}_{FV}+t\boldsymbol{v}_{gr}|b\rangle_0, \qquad (110)
\end{aligned}$$

and its square,

$$\begin{aligned}
\langle\chi|\boldsymbol{x}^2_{FV}|\chi\rangle &= -\int_{\dot{\boldsymbol{p}}\dot{\boldsymbol{q}}} d^3\dot{x}[a^*_{\dot{\boldsymbol{p}}}\bar{w}_+e^{it\omega_{\dot{p}}-i\dot{\boldsymbol{x}}\dot{\boldsymbol{p}}}+b_{\dot{\boldsymbol{p}}}\bar{w}_-e^{-it\omega_{\dot{p}}+i\dot{\boldsymbol{x}}\dot{\boldsymbol{p}}}][w_+\boldsymbol{\nabla}^2_{\dot{q}}e^{i\dot{\boldsymbol{q}}\dot{\boldsymbol{x}}}a_{\dot{\boldsymbol{q}}}e^{-it\omega_{\dot{q}}}+w_-\boldsymbol{\nabla}^2_{\dot{q}}|e^{-i\dot{\boldsymbol{q}}\dot{\boldsymbol{x}}}b^*_{\dot{\boldsymbol{q}}}e^{it\omega_{\dot{q}}}]\\
&= -\int_{\dot{\boldsymbol{q}}}[a^*_{\dot{\boldsymbol{q}}}[\boldsymbol{\nabla}^2-2it\boldsymbol{\nabla}\omega_{\dot{q}}\boldsymbol{\nabla}-it\boldsymbol{\nabla}^2\omega_{\dot{q}}-t^2(\boldsymbol{\nabla}\omega_{\dot{p}})^2]a_{\dot{\boldsymbol{q}}}\\
&\quad -b_{\dot{\boldsymbol{q}}}[\boldsymbol{\nabla}^2+2it\boldsymbol{\nabla}\omega_{\dot{p}}\boldsymbol{\nabla}+it\boldsymbol{\nabla}^2\omega_{\dot{q}}-t^2(\boldsymbol{\nabla}\omega_{\dot{q}})^2]b^*_{\dot{\boldsymbol{q}}}]\\
&= \langle a|(\boldsymbol{x}_{FV}+t\boldsymbol{v}_{gr})^2|a\rangle_0-\langle b|(\boldsymbol{x}_{FV}+t\boldsymbol{v}_{gr})^2|b\rangle_0, \qquad (111)
\end{aligned}$$

reflect the negative norm of the anti-particle modes. The spread of the relativistic state is

$$\begin{aligned}
\sigma^2_{\dot{x}}(t) &= \langle a|(\boldsymbol{x}_{FV}+t\boldsymbol{v}_{gr})^2|a\rangle_0-\langle b|(\boldsymbol{x}_{FV}+t\boldsymbol{v}_{gr})^2|b\rangle_0-(\langle a|\boldsymbol{x}_{FV}+t\boldsymbol{v}_{gr}|a\rangle_0-\langle b|\boldsymbol{x}_{FV}+t\boldsymbol{v}_{gr}|b\rangle_0)^2\\
&= \sigma^2_{\dot{x}}(0)+ta+t^2b \qquad (112)
\end{aligned}$$

with

$$\begin{aligned}
a &= (\langle a|\boldsymbol{x}_{FV}\boldsymbol{v}_{gr}+\boldsymbol{v}_{gr}\boldsymbol{x}_{FV}|a\rangle_0-2\langle a|\boldsymbol{x}_{FV}|a\rangle_0\langle a|\boldsymbol{v}_{gr}|a\rangle_0\\
&\quad -\langle b|\boldsymbol{x}_{FV}\boldsymbol{v}_{gr}+\boldsymbol{v}_{gr}\boldsymbol{x}_{FV}|b\rangle_0-2\langle b|\boldsymbol{x}_{FV}|b\rangle_0\langle b|\boldsymbol{v}_{gr}|b\rangle_0\\
&\quad +2\langle a|\boldsymbol{x}_{FV}|a\rangle_0\langle b|\boldsymbol{v}_{gr}|b\rangle_0+2\langle a|\boldsymbol{v}_{gr}|a\rangle_0\langle b|\boldsymbol{x}_{FV}|b\rangle_0,\\
b &= \langle a|\boldsymbol{v}^2_{gr}|a\rangle_0-\langle a|\boldsymbol{v}_{gr}|a\rangle^2_0-\langle b|\boldsymbol{v}^2_{gr}|b\rangle_0-\langle b|\boldsymbol{v}_{gr}|b\rangle^2_0+2\langle a|\boldsymbol{v}_{gr}|a\rangle_0\langle b|\boldsymbol{v}_{gr}|b\rangle_0. \qquad (113)
\end{aligned}$$

The wave-packet can be arbitrarily narrow, the particles and the anti-particles are treated separately and remain decoupled in the absence of an external field or interactions and the group velocity, based on the relativistic dispersion relation, is bounded by c.

The expectation value of the coordinate operator in the Klein-Gordon basis, $\hat{\boldsymbol{x}}_{KG} = \hat{\boldsymbol{x}}_{FV} - i\boldsymbol{c_p}\sigma_1$,

$$\begin{aligned}\langle\chi|\boldsymbol{x}_{KG}|\chi\rangle &= i\int_{\boldsymbol{pq}} d^3\dot{x}[a^*_{\boldsymbol{p}}\bar{w}_+ e^{it\omega_p - i\boldsymbol{x}_{FV}\boldsymbol{p}} + b_{\boldsymbol{p}}\bar{w}_- e^{-it\omega_p + i\boldsymbol{x}_{FV}\boldsymbol{p}}] \\ &\times[a_{\boldsymbol{q}}e^{-it\omega_q}(-\boldsymbol{\nabla}_q - \boldsymbol{c}_q\sigma_1)w_+ e^{i\boldsymbol{q}\boldsymbol{x}_{FV}} + b^*_{\boldsymbol{q}}e^{it\omega_q}(\boldsymbol{\nabla}_q - \boldsymbol{c}_q\sigma_1)w_- e^{-i\boldsymbol{q}\boldsymbol{x}_{FV}}], \quad (114)\end{aligned}$$

can be written as

$$\begin{aligned}\langle\chi|\boldsymbol{x}_{KG}|\chi\rangle &= \langle\chi|\boldsymbol{x}_{FV}|\chi\rangle + i\int_{\boldsymbol{q}} \boldsymbol{c}_q[b_{-\boldsymbol{q}}a_{\boldsymbol{q}}e^{-2it\omega_q} - a^*_{-\boldsymbol{q}}b^*_{\boldsymbol{q}}e^{2it\omega_q}] \\ &= \langle a|\boldsymbol{x} + t\boldsymbol{v}_{gr}|a\rangle_0 - \langle b|\boldsymbol{x} + t\boldsymbol{v}_{gr}|b\rangle_0 + 2\Re(i\langle b^*|e^{-2it\omega_{\boldsymbol{p}}}\boldsymbol{c_p}|a\rangle_0). \quad (115)\end{aligned}$$

When only positive or negative frequency modes are present then we recover eq. (110) but the interference of the particle and anti-particle modes yields Zitterbewegung, a fast oscillating term.

The problem, related to the interference of the particle and anti-particle states in the time evolution of the expectation value of the coordinate is nicely reflected in the expectation value of the velocity operator, defined in the Heisenberg representation,

$$\partial_t\boldsymbol{x} = -i[\boldsymbol{x}, H], \quad (116)$$

which is

$$\begin{aligned}\partial_t\boldsymbol{x} &= -i\left[\boldsymbol{x}, \frac{\boldsymbol{p}^2}{2m}(\sigma_3 + i\sigma_2) + m\sigma_3\right] \\ &= \frac{\boldsymbol{p}}{m}(\sigma_3 + i\sigma_2) \quad (117)\end{aligned}$$

in the Klein-Gordon basis. Its average in the state (62),

$$\begin{aligned}\langle\chi|\partial_t\hat{\boldsymbol{x}}_{KG}|\chi\rangle &= \int_{\boldsymbol{pq}} d^3x[a^*_{\boldsymbol{p}}\bar{u}_{\boldsymbol{p}}e^{i\omega_{\boldsymbol{p}}t - i\boldsymbol{p}\boldsymbol{x}} + b_{\boldsymbol{p}}\bar{v}_{\boldsymbol{p}}e^{-i\omega_{\boldsymbol{p}}t + i\boldsymbol{p}\boldsymbol{x}}] \\ &\times\frac{\boldsymbol{q}}{m}(\sigma_3 + i\sigma_2)[a_{\boldsymbol{q}}u_{\boldsymbol{q}}e^{-i\omega_{\boldsymbol{q}}t + i\boldsymbol{q}\boldsymbol{x}} + b^*_{\boldsymbol{q}}v_{\boldsymbol{q}}e^{i\omega_{\boldsymbol{q}}t - i\boldsymbol{q}\boldsymbol{x}}] \\ &= \int_{\boldsymbol{p}} \frac{\boldsymbol{p}}{m}[a^*_{\boldsymbol{p}}a_{\boldsymbol{p}}\bar{u}_{\boldsymbol{p}}(\sigma_3 + i\sigma_2)u_{\boldsymbol{p}} + b_{\boldsymbol{p}}b^*_{\boldsymbol{p}}\bar{v}_{\boldsymbol{p}}(\sigma_3 + i\sigma_2)v_{\boldsymbol{p}} \\ &-e^{2i\omega_{\boldsymbol{p}}t}a^*_{-\boldsymbol{p}}b^*_{\boldsymbol{p}}\bar{u}_{\boldsymbol{p}}(\sigma_3 + i\sigma_2)v_{\boldsymbol{p}} - e^{-2i\omega_{\boldsymbol{p}}t}b_{-\boldsymbol{p}}a_{\boldsymbol{p}}\bar{v}_{\boldsymbol{p}}(\sigma_3 + i\sigma_2)u_{\boldsymbol{p}}] \\ &= \int_{\boldsymbol{p}} \frac{\boldsymbol{p}}{\omega_p}[a^*_{\boldsymbol{p}}a_{\boldsymbol{p}} - b_{\boldsymbol{p}}b^*_{\boldsymbol{p}} - e^{2i\omega_{\boldsymbol{p}}t}a^*_{-\boldsymbol{p}}b^*_{\boldsymbol{p}} - e^{-2i\omega_{\boldsymbol{p}}t}b_{-\boldsymbol{p}}a_{\boldsymbol{p}}], \quad (118)\end{aligned}$$

shows clearly fast oscillation owing to the particle and anti-particle interference.

The advantage of the decoupled basis is the absence of the non-physical Zitterbewegung contributions. However one should be aware that the gauge transformations which are strictly local must be redefined when moving from the non-relativsitic to the relativistic domain.

C. External field

When the external electromagnetic field is is assumed then the Klein-Gordon spinor (44) is given by

$$\begin{pmatrix}\chi_1\\ \chi_2\end{pmatrix} = \frac{1}{2}\begin{pmatrix}\phi + \frac{i}{m}D_0\phi\\ \phi - \frac{i}{m}D_0\phi\end{pmatrix} \tag{119}$$

and the Klein-Gordon equation, (32), can be written as

$$iD_0\chi = H\chi \tag{120}$$

cf. (47), where

$$H = -\frac{\boldsymbol{D}^2}{2m}(\sigma_3 + i\sigma_2) + m\sigma_3. \tag{121}$$

and the covariant derivatives $D_0 = \partial_0 + ie\varphi$, $\boldsymbol{D} = \boldsymbol{\nabla} - ie\boldsymbol{A}$ are given in terms of the vector potential $A^\mu = (\varphi, \boldsymbol{A})$. The equation of motion is therefore

$$i\partial_t\chi = \left[-\frac{\boldsymbol{\nabla}^2}{2m}(\sigma_3 + i\sigma_2) + m\sigma_3 - \left(\frac{ie}{m}\boldsymbol{A}\boldsymbol{\nabla} + \frac{ie}{2m}\boldsymbol{\nabla}\boldsymbol{A}\right)(\sigma_3 + i\sigma_2) + e\phi\right]\chi. \tag{122}$$

The spatial components of the vector potential mix the particle and anti-particle modes and the states which decouple, the particle and the anti-particle states in the presence of the vector potential can not be found in a simple, closed form. When the wave function is represented by a Fourier integral as in eq. (62) then the external field pieces, the last two terms in the right hand side, mix to the plane wave with momentum $\boldsymbol{p}$ with momentum $\boldsymbol{q} \neq \boldsymbol{p}$ with an amplitude which is proportional to $A^\mu_{\pm\omega_{\boldsymbol{p}-\boldsymbol{q}},\boldsymbol{p}-\boldsymbol{q}}$. This not only reshuffles the Fourier modes as in the non-relativistic case but mixes the particle and the anti-particle modes and renders the formalism inconsistent, discussed in the following examples.

1. One dimensional potential barrier

The simplest, non-trivial example is the one dimensional stationary state of a particle in the z-direction in the presence of step potential, $U(z) = U_0\Theta(z)$, $U_0 > 0$. The wave function,

$$\psi(t, z) = \chi(z)e^{-itE}, \tag{123}$$

satisfies the Klein-Gordon equation

$$[(E - U(z))^2 + \nabla_z^2 - m^2]\chi(z) = 0. \tag{124}$$

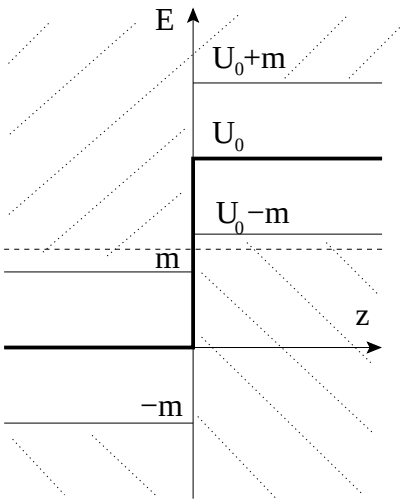

FIG. 3: A step function potential, $U(z)$, shown by fat solid line. The particle and anti-particle continuum is indicated by dotted lines. The particle and anti-particle states are mixed at the energy level, denoted by a dashed line.

The solution is of the form $\chi(z) = \Theta(-z)[\chi_i(z) + \chi_r(z)] + \Theta(z)\chi_t(z)$ where

$$\chi_i(z) = e^{ipz}, \quad \chi_r(z) = be^{-ipz}, \quad \chi_t(z) = de^{ip'z}, \tag{125}$$

with $p = \sqrt{E^2 - m^2}$ and $p' = \sqrt{(E - U_0)^2 - m^2}$. The matching conditions are the same as in the non-relativistic case, namely the continuity of the wave function,

$$1 + b = d \tag{126}$$

and its first derivative,

$$1 - b = d\xi \tag{127}$$

where $\xi = \frac{p'}{p}$. Therefore the coefficients b and d are

$$b = \frac{1 - \xi}{1 + \xi}, \quad d = \frac{2}{1 + \xi}. \tag{128}$$

Let us consider strong step potential, $U_0 > 2m$, and the energy level $m < E < U_0 - m$, indicated by dashed line in Fig. 3, where there are extended states on both sides of the singularity. The current

$$j^z = \frac{1}{2im}(\chi^* \nabla_z \chi - \nabla_z \chi^* \chi), \tag{129}$$

yields $j_i^z(0) = p$, $j_r^z(0) = -|b|^2 p$, $j_t^z(0) = |d|^2 p'$ and one finds positive reflection and transmission coefficients,

$$R = |b|^2 = \frac{(1 - \xi)^2}{(1 + \xi)^2}, \quad T = |d|^2 \xi = \frac{4\xi}{(1 + \xi)^2} \tag{130}$$

and the fulfillment of current conservation, $j_i^z(0) + j_r^z(0) = j_t^z(0)$, assures $R + T = 1$. Since $\xi > 0$ we have $0 \leq R, T \leq 1$. The mixing of the particle-anti particle modes in the state is reflected by the sign of the charge density (35), $j^0(z) = E/m > 0$ for $z < 0$ and $j^0(z) = (E - U)/m < 0$ for $z > 0$.

2. Spherical potential well

As another example we consider the Oppenheimer-Schiff-Snyder effect, displayed by a particle moving in a spherical potential, $U(u)$. The Klein-Gordon equation assumes the form

$$[(\partial_0 + iU(r))^2 - \Delta + m^2]\phi(x) = 0 \tag{131}$$

whose stationary states will be sought in the parametrization

$$\phi_{lm}(x) = \eta_\ell(r) Y_m^\ell(\theta, \phi) e^{-itE}. \tag{132}$$

The radial wave function of a given ℓ-shell satisfies the equation

$$\left[(E - U(r))^2 + \frac{1}{r^2}\partial_r r^2 \partial_r - \frac{l(l+1)}{r^2} - m^2\right] \eta_\ell(r) = 0. \tag{133}$$

It is advantageous to separate the radial integral measure by writing $\eta(r) = u(r)/r$ and

$$\left[\partial_r^2 - \frac{l(l+1)}{r^2} + (E - U(r))^2 - m^2\right] u_\ell(r) = 0. \tag{134}$$

Let us consider an attractive square well potential, $U(r) = -U_0\Theta(r - R)$, and look for the stationary states with energy $-m < E < m$ and $E > m - U_0$, allowing the mixing of the particle and anti-particle modes, cf. Fig. 4. We consider the s-wave sector, $\ell = 0$, only for the sake of simplicity where the radial wave function obeys the equation

$$u_0'' = [m^2 - (E - U(r))^2]u_0. \tag{135}$$

The solution for $r < R$ is

$$u_0 = \sin \kappa r \tag{136}$$

with $\kappa = \sqrt{(E + U_0)^2 - m^2}$, the component $\cos \kappa r$ being suppressed by the regularity of the wave function at the origin, $u_\ell(0) = 0$. In the exterior region, $r > R$, we have

$$u_0 = ae^{-kr} \tag{137}$$

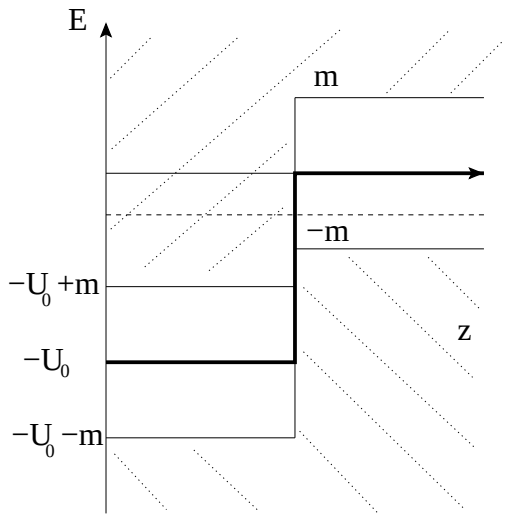

FIG. 4: A spherical potential well, $U(r)$, using the convention of Fig. 3.

with $k = \sqrt{m^2 - E^2}$. The matching conditions,

$$\sin \kappa R = ae^{-kR}, \quad \kappa \cos \kappa R = -kae^{-kR} \tag{138}$$

yield the transcendental equation

$$\tan R\kappa = -\frac{\kappa}{k}, \tag{139}$$

whose more detailed form,

$$\tan R\sqrt{(E+U_0)^2 - m^2} = -\sqrt{\frac{(E+U_0)^2 - m^2}{m^2 - E^2}}, \tag{140}$$

can be solved graphically. One can easily see an interesting qualitative feature, namely there are more and more solutions as $U_0 \to \infty$, shown qualitatively in Fig. 5. As U_0 is increases the extended particle and anti-particle modes, influenced less by the potential, decrease their energy, as expected. But some bound states are formed withing the forbidden gap, coming both from the particle and the anti-particle continuum. Note that there are values of U_0 which produce two bound states, corresponding to the same continuous curve in Fig. 5. This is the first surprise, namely the same potential can bind both particles and anti-particles. The other surprise appear as U_0 is increase until the two bound states coincide where the curve has vertical slope. The system lost a particle and an anti-particle state at this point. The energy level, given by the matching condition, (140), becomes complex at this point indicating that this is an unstable, virtual particle-anti particle pair which can penetrate up a finite distance into the forbidden, $r > R$ region.

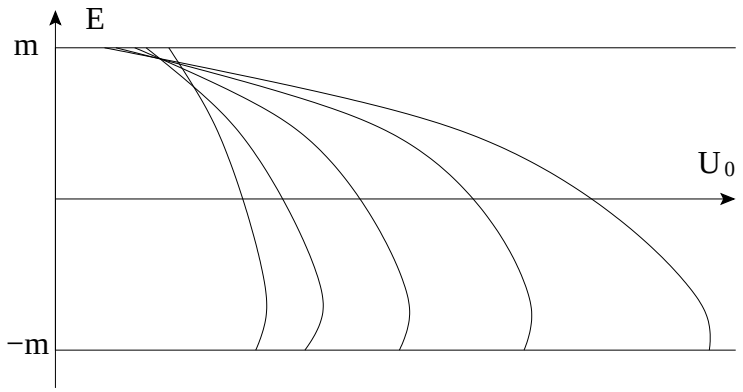

FIG. 5: A spherical potential well, $U(r)$, using the convention of Fig. 3.

III. FERMIONS

A. Heuristic derivation of Dirac equation

By following Dirac's intuitive derivation we seek an equation of motion which is of first order in the space-time derivatives by taking formally the square root of the Klein-Gordon equation (31). The result, written as

$$i\partial_0\psi = (-i\boldsymbol{\alpha}\boldsymbol{\partial} + \beta m)\psi = H\psi, \tag{141}$$

where the Hamiltonian,

$$H = \boldsymbol{\alpha}\boldsymbol{p} + \beta m, \tag{142}$$

contains the formal symbols, $\boldsymbol{\alpha}$ and β. The square of this equation should reproduce the Klein-Gordon equation,

$$\begin{aligned} -\partial_0^2\psi &= (-i\boldsymbol{\alpha}\cdot\boldsymbol{\partial} + \beta m)^2\psi \\ &= [-\{\alpha_j,\alpha_k\}\partial_j\partial_k + \beta^2m^2 - mi\{\alpha_j,\beta\} + \partial_j]\psi, \end{aligned} \tag{143}$$

where the anti-commutator $\{A,B\} = AB + BA$ has been introduced and the identity

$$\begin{aligned} \alpha_j\alpha_k\partial_j\partial_k &= \frac{1}{2}(\{\alpha_j,\alpha_k\} + [\alpha_j,\alpha_k])\frac{1}{2}(\{\partial_j,\partial_k\} + [\partial_j,\partial_k]) \\ &= \frac{1}{4}(\{\alpha_j,\alpha_k\} + [\alpha_j,\alpha_k])\{\partial_j,\partial_k\} \\ &= \frac{1}{2}\{\alpha_j,\alpha_k\}\partial_j\partial_k \end{aligned} \tag{144}$$

was used in the last equation. We arrive at the Klein-Gordon equation by requiring

$$\{\alpha_j,\alpha_k\} = 2\delta_{j,k}, \quad \beta^2 = \mathbb{1}, \quad \{\boldsymbol{\alpha},\beta\} = 0. \tag{145}$$

A covariant notation is established by introducing $\gamma^\mu = (\beta, \beta\boldsymbol{\alpha})$ and writing the Dirac equation as

$$(i\gamma^\mu \partial_\mu - m)\psi(x) = (i\gamma^0\partial_0 + i\boldsymbol{\gamma}\boldsymbol{\nabla} - m)\psi(x) = (i\partial\!\!\!/ - m)\psi(x) = 0 \tag{146}$$

where the constraints (145) are

$$\{\gamma^\mu, \gamma^\nu\} = 2g^{\mu\nu} = 2\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}. \tag{147}$$

Covariance with respect to Lorentz transformation requires that the objects γ^μ transform as contravariant four-vectors.

The "square" of Dirac equation, calculated before can now be written in a simple, covariant form,

$$(i\partial\!\!\!/ - m)(i\partial\!\!\!/ + m) = -\partial^2 - m^2. \tag{148}$$

One can show that the simplest realization of the objects $\boldsymbol{\alpha}$ and β is in terms of 4×4 matrices,

$$\beta = \gamma^0 = \begin{pmatrix} \mathbb{1} & 0 \\ 0 & -\mathbb{1} \end{pmatrix}, \qquad \boldsymbol{\alpha} = \begin{pmatrix} 0 & \boldsymbol{\sigma} \\ \boldsymbol{\sigma} & 0 \end{pmatrix}, \qquad \boldsymbol{\gamma} = \begin{pmatrix} 0 & \boldsymbol{\sigma} \\ -\boldsymbol{\sigma} & 0 \end{pmatrix}, \tag{149}$$

and any other set of 4×4, verifying the same conditions, (145), can be obtained by a unitary transformation.

The hermitian conjugate of the wave function satisfies the equation of motion

$$i\partial_\mu \psi^\dagger(x)\gamma^{\mu\dagger} + \psi(x)m = 0. \tag{150}$$

Since some of the the Dirac matrices, γ^μ, are not Hermitian, $\gamma^{0\dagger} = \gamma^0$, $\gamma^{j\dagger} = -\gamma^j$, this equation is not covariant. But it is easy to find a linear combination of the components of $\psi^\dagger$ which satisfies covariant equation of motion. The starting point is the relation

$$\gamma^0\gamma^\mu\gamma^0 = \gamma^{\mu\dagger} \tag{151}$$

which suggests that the Dirac conjugate,

$$\bar{\psi} = \psi^\dagger\gamma^0, \tag{152}$$

will satisfy covariant equation. In fact, inserting $\gamma^{\mu\dagger}$ of eq. (151) into eq. (150) we find

$$i\partial_\mu\bar{\psi}\gamma^\mu + \bar{\psi}m = 0. \tag{153}$$

But there is an important difference between the use of the Klein-Gordon and the Dirac conjugation. The former is used in the definition of the scalar product to render the Hamiltonian (48) Hermitian. The Hamiltonian (142) is Hermitian from the very beginning when the matrix $\boldsymbol{\alpha}$, given by eq. (149) is used together with the usual Hermitian conjugate in the scalar product,

$$\langle\psi|\psi'\rangle = \int d^3x \psi^\dagger(\boldsymbol{x})\psi'(\boldsymbol{x}). \tag{154}$$

The Dirac conjugate, $\bar{\psi}$, appears only as an auxiliary variable to make the equation of motion for $\psi^\dagger$ covariant.

The Dirac-equation can be derived as an Euler-Lagrange equation from the Lagrangian

$$L = \frac{i}{2}[\bar{\psi}\gamma^\mu(\partial_\mu\psi) - (\partial_\mu\bar{\psi})\gamma^\mu\psi] - m\bar{\psi}\psi. \tag{155}$$

The Noether current of the $U(1)$ symmetry, $\psi(x) \to e^{i\theta}\psi(x)$, $\bar{\psi}(x) \to e^{-i\theta}\bar{\psi}(x)$ is

$$j^\mu = \bar{\psi}\gamma^\mu\psi. \tag{156}$$

An important difference compared to the Klein-Gordon equation driven scalar particle is that the conserved current of an equation of motion which is a first order differential equation contains no derivative. As a result $j^0 = \psi^\dagger\psi$ is positive definite and no states with negative norm arise.

The derivation of the Dirac equation, presented above is simple and heuristic but leaves the physical interpretation of the Dirac space, the four dimensional linear space of the Dirac spinor, ψ_a, unclear. To have better understanding of the role played by these components we re-derive eq. (146) as the simplest equation which governs an elementary particle, equipped with relativistic symmetries.

B. Spinors

The four dimensional Dirac-space, emerging from the heuristic argument, has physically interpretable structure. To discover it we need a rather lengthy detour into the representation of the space-time symmetries, realized by spin half particles.

1. Non-relativistic spinors

Elementary systems: Elementary quantum systems are defined with respect to their symmetry properties. Suppose that we know that our system under consideration is lacking of any internal structure and displays a symmetry with respect to transformations, belonging to a group,

G. Therefore there is a linear, unitary or anti-unitary operator, $U(g)$, corresponding to each symmetry transformation which acts in the linear space of states and this representation of the group preserves the algebraic structure of group multiplication, $U(gg') = U(g)U(g')$.

What can be the consequence in this algebraic structure that our system is elementary? It is natural to expect that any state can be obtained from a fixed state, $|\psi_0\rangle$, by the application of symmetry transformations. In fact, suppose that this is not true and there is a state $|\psi'\rangle$ which is not in the linear space generated by the set of vectors $U(g)|\psi_0\rangle$, $g \in G$. Then we can safely ignore the state $|\psi'\rangle$ in the discussion of our elementary system because it plays no role in realizing the symmetry and its inaccessibility by the application of symmetry transformations should come from some internal structure. The possibility of generating all states of the system from an arbitrary but fixed state is called irreducibility. The states of an elementary system with a symmetry group G are therefore vectors of irreducible representation of G.

The construction of irreducible representation is quite different for discrete and continuous groups. This is the reason that one carefully separates these cases and considers first the connected components of symmetry groups only. For instance, spatial rotations in n-dimensions, $\boldsymbol{x} \to R\boldsymbol{x}$, are realized by $n \times n$ orthogonal matrices, $R^{\rm tr}R = \mathbb{1}$ and the determinant of this equation, $(\det R)^2 = 1$, assures $\det R = \pm 1$ because the determinant of a real matrix is real. The determinant is a continuous function of the matrix elements, hence the group $O(n)$ has two disconnected components, $O_\pm(n) = \{R | R^{\rm tr}R = \mathbb{1},\ \det R = \pm 1\}$. The subset $O_+(n) = SO(n)$ contains the identity and is a subgroup. It is easy to establish a bijective relation between $O_+(n)$ and $O_-(n)$, it is given by the inversion of a coordinate, $P_1 : (x_1, x_2, \ldots, x_n) \to (-x_1, x_2, \ldots, x_n)$ or spatial inversion, $P : \boldsymbol{x} \to -\boldsymbol{x}$. In fact, we have $O_-(n) = P_1 O_+(n)$ and $O_-(n) = P_1 O_+(n)$ in even or odd dimensions, respectively. One works out the irreducible representation for $SO(n)$ first and extends them over $O(n)$ in the second step.

Irreps of $SO(3)$: It is known that the irreducible representations of the rotation group $G = SO(3)$ are given by the rotational multiplets,

$$\mathcal{H}_J = \left\{ \sum_{m=-J}^{J} c_m |J, m\rangle \right\}, \tag{157}$$

of dimension $2J + 1$ where J is integer, $J = 0, 1, \ldots$, or half-integer, $J = \frac{1}{2}, \frac{3}{2}, \ldots$. Thus the states of elementary systems with a rotational degree of freedom can be represented by the linear superposition of basis vectors $\{|J, m\rangle\}$ for some J. Any representation can be written as the direct sum of irreducible representations, justifying our definition of elementary system by means of its symmetry properties.

Tensors and spinors are the wave functions belonging to states with integer and half-integer angular momentum J, respectively. The difference between them is their response to rotations by 2π. Let us denote the matrix performing a rotation by angle α around the axis $\boldsymbol{n}$ by $R_{\boldsymbol{n}}(\alpha)$. We have

$$U(R_{\boldsymbol{n}}(\alpha)) = e^{-i\alpha \boldsymbol{n}\boldsymbol{L}} \tag{158}$$

where $\boldsymbol{L}$ is the angular momentum operator and the Wigner matrix elements,

$$\mathcal{D}^{(J)}_{m,m'}(R_{\boldsymbol{n}}(\alpha)) = \langle J, m|e^{-i\alpha \boldsymbol{n}\boldsymbol{L}}|J, m'\rangle, \tag{159}$$

satisfy the equation

$$\mathcal{D}^{(J)}_{m,m'}(R_{\boldsymbol{n}}(2\pi)) = \delta_{m,m'}e^{-2\pi i m} = \begin{cases} +1 & J = \text{integer}, \\ -1 & J = \text{half integer}. \end{cases} \tag{160}$$

The simplest non-trivial, tensor and vector representations belong to $J = \frac{1}{2}$ and 1, respectively. $\mathcal{H}_1$ is span by the three-vector $\boldsymbol{x}$ and vectors $\psi = (\psi_1, \psi_2)$ of the two dimensional $\mathcal{H}_{1/2}$ are called rotational spinors.

Fundamental representation of $SU(2)$: We first show that the representations of the group $SU(2)$ can be considered as the representations of the rotation group, $SO(3)$. This circumstance provides us a simple way to generalize the non-relativistic spinors, defined by the rotation group, for relativistic spinors, corresponding to the Lorentz group. Let us consider a linear combination of Pauli matrices and the identity,

$$A(a, \boldsymbol{a}) = a\mathbb{1} + i\boldsymbol{a}\boldsymbol{\sigma} = \begin{pmatrix} a + ia_3 & ia_1 + a_2 \\ ia_1 - a_2 & a - ia_3 \end{pmatrix}, \tag{161}$$

where a and $\boldsymbol{a}$ are real numbers, constrained by the condition

$$\det A(a, \boldsymbol{a}) = a^2 + \boldsymbol{a}^2 = 1. \tag{162}$$

Such a matrix structure is preserved under multiplication. In fact, identity

$$\sigma_a\sigma_b = \mathbb{1}\delta_{ab} + i\epsilon_{abc}\sigma_c \tag{163}$$

can be used to write the product of two such matrices as

$$A(a, \boldsymbol{a})A(b, \boldsymbol{b}) = A(ab - \boldsymbol{a}\boldsymbol{b}, a\boldsymbol{b} + b\boldsymbol{a} - \boldsymbol{a} \times \boldsymbol{b}). \tag{164}$$

Furthermore, this result shows that $A^{-1}(a,\boldsymbol{a}) = A(a,-\boldsymbol{a}) = A^{\dagger}(a,\boldsymbol{a})$ hence these matrices form the group $SU(2)$. Another parametrization of the $SU(2)$ matrices, better known in quantum mechanics, is

$$A_{\boldsymbol{n}}(\alpha) = e^{-\frac{i}{2}\alpha\boldsymbol{n}\boldsymbol{\sigma}} = \mathbb{1}\cos\frac{\alpha}{2} - i\boldsymbol{n}\boldsymbol{\sigma}\sin\frac{\alpha}{2}, \tag{165}$$

with $\boldsymbol{n}^2 = 1$.

The non-relativistic $SU(2)$ spinor is a two component quantity, ψ_a, $a = 1,2$, transforming according to the fundamental representation of the group $SU(2)$, $\psi \to A\psi$, and can be interpreted as a wave function. Having two components, the state, represented by this wave function should have spin $s = 1/2$. The complex conjugate of this transformation rule, $\eta \to A^*\eta$, yields another two dimensional representation. But this is unitary equivalent with the fundamental representation. Two representations, $U(g)$ and $U'(g)$, are unitary equivalent if there is g-independent unitary operator, V, which brings it into the other, $U'(g) = V^{\dagger}U(g)V$. In fact, the equation

$$(i\boldsymbol{\sigma})^* = \sigma_2 i\boldsymbol{\sigma}\sigma_2 = \sigma_2^{\dagger} i\boldsymbol{\sigma}\sigma_2 \tag{166}$$

can be used to establish $A^* = \sigma_2 A \sigma_2$. This equation together with eq. (163) define the Pauli matrices.

Adjoint, vector representation of $SU(2)$: The simplest higher dimensional, so called adjoint representation consists of two index spinors, X_{ab}, $a,b = 1,2$, considered as a matrix, transforming as

$$X \to AXA^{\dagger}, \tag{167}$$

indicating that the left and the right indices transform according to the fundamental representation and its complex conjugate, respectively. This representation is 8 dimensional but note that if X is Hermitian then this property is preserved. Thus it is advantageous to focus our attention to Hermitian matrices which can be written as

$$X(x^{\mu}) = x^0\mathbb{1} + \boldsymbol{x}\boldsymbol{\sigma} = \begin{pmatrix} x^0 + x^3 & x^1 - ix^2 \\ x^1 + ix^2 & x^0 - x^3 \end{pmatrix}, \tag{168}$$

with real coefficients x^{μ}. Note that $\det X$ is preserved during this transformation. This matrix can be considered as the spin wave function of two spin half particles where we use a basis which is transformed by the unitary matrix σ_2 for the particle of the right index. The rule $\frac{1}{2}\otimes\frac{1}{2} = 0\oplus 1$ of the addition of angular momentum suggests that we have a spin zero singlet and a spin one triplet

subspaces. It is clear that the former belongs to the component x^0, being invariant under $SU(2)$ transformations.

Relation between $SO(3)$ and $SU(2)$: Now we show that the triplet multiplet is realized by the components $(x^1, x^2, x^3) = \boldsymbol{x}$. The length of the vector $\boldsymbol{x}$, being related to the determinant, $\det X = -\boldsymbol{x}^2$, is preserved by the $SU(2)$ transformation, given by (167). We therefore have found a linear, length preserving transformation of three-vectors for each $SU(2)$ group elements. The group $SU(2)$ has the same dimension as $SO(3)$ which makes it possible that the different $SU(2)$ transformations cover the group $SO(3)$. To prove that this possibility is indeed realized it is sufficient to consider infinitesimal $SU(2)$ transformation using the parametrization (165),

$$\begin{aligned} A_{\boldsymbol{n}}(\alpha) X(0,\boldsymbol{x}) A^{\dagger}_{\boldsymbol{n}}(\alpha) &\approx \left(\mathbb{1} - i\frac{\alpha}{2}\boldsymbol{n}\boldsymbol{\sigma}\right) \boldsymbol{x}\boldsymbol{\sigma} \left(\mathbb{1} + i\frac{\alpha}{2}\boldsymbol{n}\boldsymbol{\sigma}\right) \\ &\approx \boldsymbol{x}\boldsymbol{\sigma} - i\frac{\alpha}{2}[\boldsymbol{n}\boldsymbol{\sigma}, \boldsymbol{x}\boldsymbol{\sigma}] \\ &= X(0, \boldsymbol{x} + \alpha \boldsymbol{n} \times \boldsymbol{x}), \end{aligned} \tag{169}$$

which is indeed an infinitesimal rotation. The repeated application of this argument establishes the result for finite α.

Topology: We have just proven is that the transformation (167) realizes a mapping, $SU(2) \to SO(3)$, $A \to R(A)$. The image is not the full $O(3)$ group because the repetition of infinitesimal rotations can cover the connected subgroup, $SO(3)$ only.

Our mapping $SU(2) \to SO(3)$ is a two-to-one since $R(-A) = R(A)$. Though the group manifold $SU(2)$ is simply connected this mapping shows that the group $SO(3)$ has multiply connected topology: In fact, the diametrically opposite $SO(3)$ rotations, $R_{\boldsymbol{n}}(\pi)$ and $R_{\boldsymbol{n}}(-\pi)$, are identical because one is obtained from the other by rotation by 2π, a symmetry transformation of vectors. Hence a closed loop in the $SO(3)$ group manifold which starts at $\mathbb{1}$, passes to $R_{\boldsymbol{n}}(\pi)$, "jumps" to $R_{\boldsymbol{n}}(-\pi)$ and returns back to $\mathbb{1}$ can not be deformed in a continuous manner into a closed loop passing by $\mathbb{1}$ and not using the relation $R_{\boldsymbol{n}}(\pi) = R_{\boldsymbol{n}}(-\pi)$, cf. Fig. 6. On the contrary, any closed loop can be deformed continuously to a single point in $SU(2)$ which has the topology of a three-sphere, cf. eq. (162). The spinors transforms as $\psi \to -\psi$ under rotation by 2π and they are preserved by rotation $\alpha = 4\pi$.

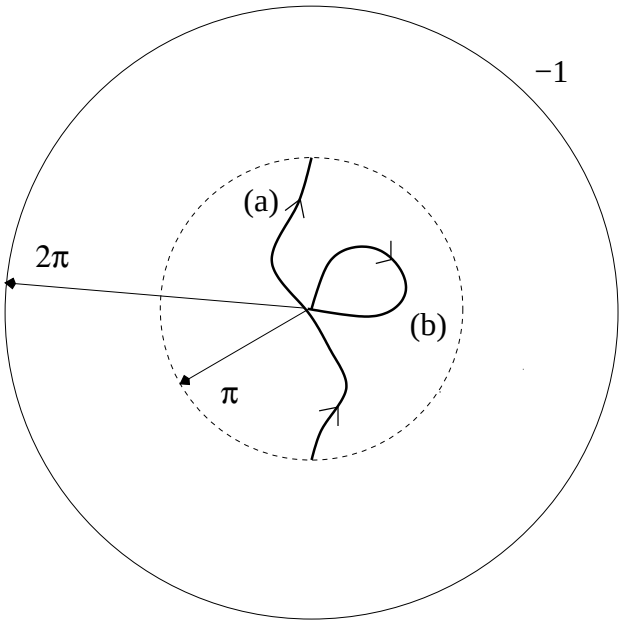

FIG. 6: The group space of $SU(2)$ consists of a sphere or radius 2π in the space $\alpha\boldsymbol{n}$, the surface being identified with $-\mathbb{1}$. The group $SO(3)$ is within the sphere of radius π. Two closed loops representing the two homotopy classes of the group $SO(3)$ are shown, starting and ending at the identity, located at the origin. The loop (a) reaches $R_{\boldsymbol{n}}(\pi)$, jumps to $R_{\boldsymbol{n}}(-\pi)$. The loop (b) does not use the identity of the diametrically opposite points. The (b) is homotopic with the trivial loop, resting at $\mathbb{1}$ and the (a) is an element of the only non-trivial homotopy class.

2. Projective representations

The vectors $|\psi\rangle$ and $e^{i\alpha}|\psi\rangle$ represent the same physical state hence the representation,$U(g)$, is to be generalized to a projective representation,

$$U(gg') = U(g)U(g')e^{i\alpha(g,g')}, \tag{170}$$

where the function $\alpha(g, g')$ is arbitrary, except that group multiplication should remain associative, $\alpha(g_3, g_2g_1) + \alpha(g_2, g_1) = \alpha(g_3, g_2) + \alpha(g_3g_2, g_1)$. The obvious question which arises at this point is whether this phase can be eliminated by a "gauge transformation", $U(g) \to U(g)e^{i\beta(g)}$. There are two conditions to satisfy if we want to reduce a projective representation of a connected group to a non-projective representation:

1. Local: One can eliminate the phase factor locally, in the vicinity of the identity if the group has no central charge (a group has a central charge if the commutator of some generators contain a term, proportional to the unit operator). It is worth to mention that the central charge

of a semi-simple group (a group is semi-simple if has no generators which commute with all the other generators) can be eliminated by the appropriate redefinition of the generators.

2. Global: The phase factor can be eliminated globally, in the whole group, if in addition the topology of the group is simply connected. In case of a multiply connected topology the phase $\alpha(g, g')$ gives a representation of the fundamental group (the group consisting of the connected components of closed loops within the group).

Let us apply these theorems for the rotation group, $SO(3)$, which is semi-simple but doubly connected, its fundamental group $\pi_1(SO(3)) = Z_2$. The spin is pseudo-scalar and is left unchanged by space inversion therefore $U(P)$ is a unitary, diagonal 2×2 matrix. Two inversions, executed subsequently restores the original state therefore one expects $U^2(P) = \mathbb{1} \in Z_2$. But there are projective representations with $U^2(P) = -\mathbb{1} \in Z_2$, as well. In this case two inversions amounts to a rotation by 2π, an invisible transformation in classical physics. We have therefore either $U(P) = \pm\mathbb{1}$ ($U^2(P) = \mathbb{1}$) or $U(P) = \pm i\mathbb{1}$ ($U^2(P) = -\mathbb{1}$).

Every rotation by 2π is represented by the same transformation, $R_{\boldsymbol{n}}(2\pi) = \mathbb{1}$ in $SO(3)$ and $A_{\boldsymbol{n}}(2\pi) = -\mathbb{1}$ in $SU(2)$. The family of rotations with $0 < \alpha < 2\pi$ generate a topological non-trivial closed loop in the group space. Therefore $U(R_{\boldsymbol{n}}(2\pi)) = \mathbb{1}e^{i\theta}$, there the phase factor gives a representation of the fundamental group, $\pi_1(SO(3)) = Z_2$. This restricts the three-dimensional particles to being bosons ($\theta = 0$) or fermions ($\theta = \pi$). The fundamental group is two dimensional rotations is larger, $\pi_1(SO(2)) = Z$, the additive group of integers, each homotopy class is characterized by the winding number of the loops. As a result the phase θ is arbitrary in two dimensions, the two dimensional edge states of the quantum Hall effects are anyons.

3. Relativistic spinors

Fundamental representation of $SL(2c)$: It is not difficult to find spinors for the Lorentz group along the line of though of the non-relativistic construction. One starts with $SL(2, c)$ groups, consisting of complex 2×2 matrices with unit determinant and its fundamental representation realized by $SL(2, c)$ spinors, as $\psi \to A\psi$. The parametrization (161), subject of the constraint (162) remains valid for $SL(2, c)$ matrices except that a and $\boldsymbol{a}$ are complex numbers. Therefore $SL(2, c)$ is a six dimensional group. The generalization of the parametrization (165) is obtained by letting α complex.

An important difference with respect to the non-relativistic, $SU(2)$ spinors is that the fun-

damental representation and its complex conjugate are not unitary equivalent anymore because (166) is insufficient to establish the unitary equivalence for complex α. We follow van der Waerden conventions for $SL(2,c)$ spinors and use the notation $a = 1, 2$ and $\dot{a} = 1, 2$ for the indices of the fundamental and the complex conjugate spinor representation where the $SL(2,c)$ transformation rules are

$$\xi_a \to A_a{}^b \xi_b, \quad \eta_{\dot{a}} \to A^{*\dot{b}}_{\dot{a}} \psi_{\dot{b}}, \tag{171}$$

Furthermore, there are covariant and contravariant spinors because

$$g_{ab} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} = i\sigma_2 = -g^{ab} \tag{172}$$

plays the role of the metric tensor since the scalar product remains invariant under $SL(2c)$ transformation,

$$\xi_a \chi^a = \xi_a g^{ab} \chi_b = \xi_1 \chi_2 - \xi_2 \chi_1 \to (\xi_a \chi_b - \xi_b \chi_a) A_1{}^a A_2{}^b = (\xi_1 \chi_2 - \xi_1 \chi_2) \det A_a{}^b = \xi_a \chi^a. \tag{173}$$

The metric tensor is the same for the spinors and complex conjugate spinors.

Lorentz group and $SL(2c)$: Let us now consider the adjoint representation of $SL(2,c)$, obeying the transformation rule of (167) where the matrix A is now belong to $SL(2,c)$ and the matrix (168) has the indices $X^{a\dot{b}}(x^\mu)$. Another version of this matrix is

$$X_{a\dot{a}} = g_{ab} g_{\dot{a}\dot{b}} X^{b\dot{b}} = i\sigma_2 (x^0 \mathbb{1} + \boldsymbol{x}\boldsymbol{\sigma}) i\sigma_2^{\rm tr} = x^0 \mathbb{1} - \boldsymbol{x}\boldsymbol{\sigma}^{\rm tr}. \tag{174}$$

The preservation of $\det X(x) = x^2$ makes it plausible that the adjoint representation, (167), is a Lorentz transformation of x^μ. This is certainly the case for the subgroup $SU(2) \subset SL(2,c)$ has we have seen in the previous Chapter. The remaining, other three dimensions of $SL(2,c)$ belong to imaginary α of the parametrization (165),

$$A_{\boldsymbol{n}}(i\beta) = e^{\frac{\beta}{2}\boldsymbol{n}\boldsymbol{\sigma}} = \mathbb{1} \cosh \frac{\beta}{2} + \boldsymbol{n}\boldsymbol{\sigma} \sinh \frac{\beta}{2}, \tag{175}$$

and the transformation with infinitesimal β,

$$\begin{aligned} A_{\boldsymbol{n}}(i\beta) X(x^0, \boldsymbol{x}) A^\dagger_{\boldsymbol{n}}(i\beta) &\approx \left(\mathbb{1} + \frac{\beta}{2}\boldsymbol{n}\boldsymbol{\sigma}\right)(x^0 \mathbb{1} + \boldsymbol{x}\boldsymbol{\sigma})\left(\mathbb{1} + \frac{\beta}{2}\boldsymbol{n}\boldsymbol{\sigma}\right) \\ &\approx x^0 \mathbb{1} + (\boldsymbol{x} + x^0 \beta \boldsymbol{n})\boldsymbol{\sigma} + \frac{\beta}{2}\{\boldsymbol{n}\boldsymbol{\sigma}, \boldsymbol{x}\boldsymbol{\sigma}\} \\ &= X(x^0 + \beta \boldsymbol{n}\boldsymbol{x}, \boldsymbol{x} + x^0 \beta \boldsymbol{n}), \end{aligned} \tag{176}$$

indeed represents a Lorentz boost with infinitesimal velocity $\boldsymbol{v} = c\beta\boldsymbol{n}$. It is easy to see that neither time nor space inversion can be generated in this manner. Thereby we obtained a two-to-one

correspondence between $SL(2,c)$ and the proper Lorentz group, $L_+^\uparrow$ and the simplest, spin half elementary objects are described by spinors ξ^a or $\eta^{\dot{a}}$.

Topology: The topology of $SU(2) = \{A_{\boldsymbol{n}}(\alpha)\}$ is S_3, the Lorentz boosts manifold, $\{A_{\boldsymbol{n}}(-i\beta)\}$, is topologically equivalent with R^3 hence the manifold $SL(2,c)$ has the same topology than $S_3 \otimes R^3$, i.e simply connected. The $SL(2,c)$ matrices A and $-A$ have identical effect in the adjoint representation hence the topology of the proper Lorentz group is double connected, $\pi_1(L_+^\uparrow) = Z_2$.

1. P: As far as P^2 is concerned we have the same possibilities as in the non-relativistic case, $P^2 = \pm\mathbb{1}$. But the representation of P is rendered more involved than for non-relativistic rotations because the Lorentz boost, $L_{\boldsymbol{v}}$, contains the spatial components of a four-vector which itself is changed by space inversion. The result, $PL(\boldsymbol{v}) = L(-\boldsymbol{v})P \neq L(\boldsymbol{v})P$, implies that space inversion can not be represented by a multiplication of the relativistic spinor components with a common phase factor as in the non-relativistic case. Therefore the irreducible representation of $L_+^\uparrow \cup L_-^\uparrow$ is constructed by means of the direct sum of two irreducible representations of $L_+^\uparrow$. The simplest possibility is to take two spin half representations, consisting of the bi-spinors

$$\psi = \begin{pmatrix} \xi^a \\ \eta_{\dot{a}} \end{pmatrix}, \tag{177}$$

   as a four dimensional reducible representation of the proper Lorentz group. One uses the fundamental representation and its complex conjugate to be able to introduce later the matrix $X^{a\dot{a}}(p^\mu)$, representing the four-momentum in the equation of motion. Another advantage of this choice, to be demonstrated in Section III B 4, is that the anti-particle of ξ (η) can be represented by η (ξ).

   The space inversion is represented by the transformation

$$P\xi^a = z_P\eta_a, \quad P\eta_a = z_P\xi^a. \tag{178}$$

   on bi-spinors, where $z_P = \pm 1$ ($P^2 = \mathbb{1}$) or $\pm i$ ($P^2 = -\mathbb{1}$). The sign of z_P drops out from expectation values but the sign of P^2 leads to different transformation rules for complex conjugate bi-spinors and influences the construction of real bi-spinors. But as long as we do not want to use our scheme to describe neutral fermions we can ignore this sign.

2. T: The time inversion does not commute with Lorentz boosts in a manner similar to the space inversion but we can represent it within the same spinor representation of the proper

Lorentz group because T is an anti-unitary operation which changes the sign of the spin. It should not mix the particle with anti-particle thus we take

$$T\xi^a = z_T g_{ab}\xi^{*b}, \quad T\eta_a = z_T g^{ab}\eta^*_b, \tag{179}$$

the multiplication with the spinorial metric tensor being the only covariant way of flipping the spin and $z_T = \pm i$ to satisfy

$$T^2 = U(R_{\boldsymbol{n}}(2\pi)) = -\mathbb{1}. \tag{180}$$

3. C: The charge conjugation exchanges particles and anti-particles and flips the spin, therefore we use the representation

$$C\xi^a = z_C g^{ab}\eta^*_a, \quad C\eta_a = -z_C g_{ab}\xi^{*b} \tag{181}$$

where $z_C = \pm 1$ or $\pm i$.

The convention, usually followed in a PTC-symmetric quantum field theory, is $z_P = -z_T = -z_C = i$.

4. Dirac equation

The ingredients of an equation of motion for a free particle are the bi-spinor $\psi = (\xi^a, \eta_{\dot{a}})$ and the energy-momentum four-vector $p^{a\dot{a}} = p^0\mathbb{1} + \boldsymbol{p\sigma}$. The only possibility to combine these element in a linear fashion is

$$\begin{aligned} p^{a\dot{a}}\eta_{\dot{a}} &= m\xi^a, \\ p_{a\dot{a}}\xi^a &= m\eta_{\dot{a}}, \end{aligned} \tag{182}$$

where m is a constant of mass dimension. One could use different masses in the two equation but by an appropriate rescaling of the spinors the two masses become identical. Note that the mass serves as a coupling constant between the two spinors which decouple in the massless limit where we can have a covariant equation of motion for one spinor only. the set of equations (182) can be written in a more explicit manner as

$$\begin{aligned} (p^0 + \boldsymbol{p\sigma})\eta &= m\xi, \\ (p^0 - \boldsymbol{p\sigma})\xi &= m\eta. \end{aligned} \tag{183}$$

By the elimination of one spinor one arrives at the second order equations

$$(p^2 - m^2)\xi = (p^2 - m^2)\eta = 0. \tag{184}$$

The first order equation, eq.(182), can be brought into a simpler form of eq. (146) by $p_\mu = i\partial_\mu$ and the introduction of the matrices

$$\gamma^0_{ch} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \gamma^j_{ch} = \begin{pmatrix} 0 & -\sigma_j \\ \sigma_j & 0 \end{pmatrix} \tag{185}$$

which are related to the "standard" Dirac matrices, given by eq. (149) by the basis transformation

$$\gamma^\mu_{ch} = U\gamma^\mu U^\dagger, \quad U = \frac{1}{\sqrt{2}}(1 - \gamma^5\gamma^0) = \frac{1}{\sqrt{2}} \begin{pmatrix} \mathbb{1} & -\mathbb{1} \\ \mathbb{1} & \mathbb{1} \end{pmatrix}, \tag{186}$$

where

$$\gamma^5 = i\gamma^0\gamma^1\gamma^2\gamma^3, \tag{187}$$

assuming the form

$$\gamma^5 = \begin{pmatrix} \mathbb{1} & 0 \\ 0 & -\mathbb{1} \end{pmatrix} \tag{188}$$

in chiral representation.

In the presence of an external vector potential, A_μ with minimal coupling, $p_\mu \to p_\mu - eA_\mu$, $\partial_\mu \to \partial_\mu + ieA_\mu$, the Dirac equation is

$$(i\not{\partial} - e\not{A} - m)\psi = 0 \tag{189}$$

and the discrete inversions, identified by requiring that the transformed equation of motion remains covariant, assume the following form:

- $P\psi = U_P\psi_P$, $f_P(t, \boldsymbol{x}) = f(t, -\boldsymbol{x})$, $A_{\mu P}(t, \boldsymbol{x}) = (\phi(t, \boldsymbol{x}), \boldsymbol{A}(t, \boldsymbol{x}))_P = (\phi(t, -\boldsymbol{x}), -\boldsymbol{A}(t, -\boldsymbol{x}))$, $U_P = i\gamma^0$, $U_P\gamma^0 U_p^{-1} = \gamma^0$, $U_P\boldsymbol{\gamma}U_P^{-1} = -\boldsymbol{\gamma}$, $(i\not{\partial} - e\not{A}_P - m)\psi_P = 0$,
- $T\psi = U_T\bar{\psi}_T$, $f_T(t, \boldsymbol{x}) = f(-t, \boldsymbol{x})$, $A_{\mu T}(t, \boldsymbol{x}) = (\phi(t, \boldsymbol{x}), \boldsymbol{A}(t, \boldsymbol{x}))_P = (-\phi(-t, \boldsymbol{x}), \boldsymbol{A}(-t, \boldsymbol{x}))$, $U_T = -i\gamma^1\gamma^3\gamma^0$, $U_T\gamma^0 U_T^{-1} = \gamma^{\mathrm{tr}0}$, $U_T\boldsymbol{\gamma}U_T^{-1} = -\boldsymbol{\gamma}^{\mathrm{tr}}$, $(i\not{\partial} - e\not{A}_T - m)\psi_T = 0$,
- $C\psi = U_C\bar{\psi}$, $U_C = -i\gamma^2\gamma^0$, $U_C\gamma^\mu U_C^{-1} = -\gamma^{\mathrm{tr}\mu}$, $(i\not{\partial} + e\not{A} - m)\psi_C = 0$.

We turn our attention to the transformation properties of the bi-spinor under proper Lorentz transformations. Let us use the parametrization (161) for $SL(2,c)$ matrices with $a, \boldsymbol{a}$ being complex numbers and write the transformation of the ξ spinor as $\xi \to A(a,\boldsymbol{a})\xi$. The complex conjugation,

$$A^*(a,\boldsymbol{a}) = \sigma_2 A(a^*,\boldsymbol{a}^*)\sigma_2, \tag{190}$$

cf. (166), and the displacement of the van der Waerden indices,

$$A_a{}^b = g_{aa'} g^{bb'} A^{a'}_{\ b'}, \tag{191}$$

with $g = i\sigma_2$ yields

$$\eta \to i\sigma_2 A^*(a,\boldsymbol{a}) i\sigma_2^{\rm tr}\eta = A(a^*,\boldsymbol{a}^*)\eta. \tag{192}$$

Hence the proper Lorentz transformations, $A_{\boldsymbol{n}}(\alpha + i\beta)$, can be parametrized by two three-vectors, $\boldsymbol{u} = \alpha\boldsymbol{n}$ and $\boldsymbol{v} = \beta\boldsymbol{n}$ which can be rearranged in an anti-symmetric tensor,

$$\omega_{\mu\nu} = \begin{pmatrix} 0 & v_1 & v_2 & v_3 \\ -v_1 & 0 & u_3 & -u_2 \\ -v_2 & -u_3 & 0 & u_1 \\ -v_3 & u_2 & -u_1 & 0 \end{pmatrix} \tag{193}$$

in a Lorentz covariant manner in manner, similar to the field strength tensor of electrodynamics. Lorentz transformation are presented on bi-spinors as

$$\psi \to e^{-\frac{i}{4}\omega_{\mu\nu}\sigma^{\mu\nu}}\psi, \tag{194}$$

where

$$\sigma^{0j} = i\begin{pmatrix} \sigma_j & 0 \\ 0 & -\sigma_j \end{pmatrix}, \quad \sigma^{jk} = \epsilon^{jk\ell}\begin{pmatrix} \sigma_\ell & 0 \\ 0 & \sigma_\ell \end{pmatrix} \tag{195}$$

in chiral representation. A representation independent form of the generators is

$$\sigma^{\mu\nu} = \frac{i}{2}[\gamma^\mu, \gamma^\nu]. \tag{196}$$

The generator of $SO(3)$ rotations,

$$S_j = \frac{1}{2}\epsilon_{jk\ell}\sigma_{k\ell}, \tag{197}$$

gives the representation of the spin in terms of Dirac matrices.

It is important to know the transformation properties of bilinear expressions of bi-spinors, $\bar\psi_a\psi_b$ which enter in expectation values. The properties of the $4\times 4 = 16$ components are listed in Table I.

TABLE I: Transformation properties of bilinears.

Bilinear	Special Lorentz tr.	Space inv.	Time inv.	Charge conj.
$S = \bar{\psi}\psi$	S	S	S	S
$P = \bar{\psi}\gamma^5\psi$	P	$-P$	P	P
$V^\mu = \bar{\psi}\gamma^\mu\psi = (V^0, \boldsymbol{V})$	$\omega^\mu{}_\nu V^\nu$	$(V^0, -\boldsymbol{V})$	$(-V^0, \boldsymbol{V})$	V^μ
$A^\mu = \bar{\psi}\gamma^5\gamma^\mu\psi$	$\omega^\mu{}_\nu V^\nu$	$(-V^0, \boldsymbol{V})$	$(-V^0, \boldsymbol{V})$	A^μ
$T^{\mu\nu} = \bar{\psi}\sigma^{\mu\nu}\psi = T^{\mu\nu}(\boldsymbol{u}, \boldsymbol{v})$	$\omega^\mu{}_\nu \omega^{\mu'}{}_{\nu'} T^{\nu\nu'}$	$T^{\mu\nu}(-\boldsymbol{u}, \boldsymbol{v})$	$T^{\mu\nu}(-\boldsymbol{u}, \boldsymbol{v})$	$T^{\mu\nu}(\boldsymbol{u}, \boldsymbol{v})$

C. Free particles

After having clarified the physical structure of the Dirac spinor in terms of the representations of the Lorentz group we turn to the simplest problem, that of a free particle.

1. Plane wave solutions

Let us now consider the plane waves solutions,

$$\psi^{(+)}(x) = e^{-ipx} u_{\boldsymbol{p}}, \qquad \psi_p^{(-)}(x) = e^{ipx} v_{\boldsymbol{p}} \tag{198}$$

where $p^0 = \omega_{\boldsymbol{p}} \geq 0$ and the bi-spinors $u_{\boldsymbol{p}}$ and $v_{\boldsymbol{p}}$ satisfy the equations

$$(\not{p} - m) u_{\boldsymbol{p}} = (\not{p} + m) v_{\boldsymbol{p}} = 0 \tag{199}$$

with $p^2 = m^2 c^2$ due to Eq. (148). The construction of the spinors $u_{\boldsymbol{p}}$ and $v_{\boldsymbol{p}}$ starts in their rest frame, $p_0^\mu = (mc, \boldsymbol{0})$ $(m^2 > 0)$,

$$(\gamma^0 - 1) u_{\boldsymbol{0}} = (\gamma^0 + 1) v_{\boldsymbol{0}} = 0, \tag{200}$$

in particular

$$\begin{aligned} u_{\boldsymbol{0}}^{(1)} &= \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} \phi^{(1)} \\ 0 \end{pmatrix}, \quad u_{\boldsymbol{0}}^{(2)} = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} \phi^{(2)} \\ 0 \end{pmatrix}, \\ v_{\boldsymbol{0}}^{(1)} &= \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ \chi^{(1)} \end{pmatrix}, \quad v_{\boldsymbol{0}}^{(1)} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ \chi^{(2)} \end{pmatrix}. \end{aligned} \tag{201}$$

The bi-spinors corresponding to an arbitrary energy-momentum p^μ on the mass shell, $p^2 = m^2c^2$ are given by

$$\begin{aligned} u_{\boldsymbol{p}}^{(\alpha)} &= \frac{\not{p}+m}{\sqrt{2m(m+\omega_p)}} u_{\boldsymbol{0}}^{(\alpha)} = \begin{pmatrix} \sqrt{\frac{m+\omega_{\boldsymbol{p}}}{2m}}\phi^{(\alpha)} \\ \frac{\boldsymbol{\sigma p}}{\sqrt{2m(m+\omega_{\boldsymbol{p}})}}\phi^{(\alpha)} \end{pmatrix} \\ v_{\boldsymbol{p}}^{(\alpha)} &= \frac{-\not{p}+m}{\sqrt{2m(m+\omega_p)}} v_{\boldsymbol{0}}^{(\alpha)} = \begin{pmatrix} \frac{\boldsymbol{\sigma p}}{\sqrt{2m(m+\omega_{\boldsymbol{p}})}}\chi^{(\alpha)} \\ \sqrt{\frac{m+\omega_{\boldsymbol{p}}}{2m}}\chi^{(\alpha)} \end{pmatrix} \end{aligned} \tag{202}$$

according to Eq. (148) where $\phi^{(\alpha)}$ and $\chi^{(\alpha)}$ are $\boldsymbol{p}$-independent, two-component spinors. The normalization is chosen in such a manner that the relations

$$\bar{u}_{\boldsymbol{p}}^{(\alpha)} u_{\boldsymbol{p}}^{(\beta)} = -\bar{v}_{\boldsymbol{p}}^{(\alpha)} v_{\boldsymbol{p}}^{(\beta)} = \delta_{\alpha,\beta}, \qquad \bar{u}_{\boldsymbol{p}}^{(\alpha)} v_{\boldsymbol{p}}^{(\beta)} = \bar{v}_{\boldsymbol{p}}^{(\alpha)} u_{\boldsymbol{p}}^{(\beta)} = 0 \tag{203}$$

hold. It is easy to see that the current (156) assumes the form (40) for plane waves. Since $j^0 = \bar{\psi}\gamma^0\psi = \psi^\dagger\psi \geq 0$ the density is non-negative in contrast to the scalar particle.

We present for later use the projection operator corresponding to the positive and negative energy bi-spinors,

$$P_+(\boldsymbol{p}) = \sum_{\alpha=1}^{2} u_{\boldsymbol{p}}^{(\alpha)} \otimes \bar{u}_{\boldsymbol{p}}^{(\alpha)} = \frac{\not{p}+m}{\sqrt{2m(m+\omega_{\boldsymbol{p}})}} \frac{1+\gamma^0}{2} \frac{\not{p}+m}{\sqrt{2m(m+\omega_{\boldsymbol{p}})}} = \frac{\not{p}+m}{2m} \tag{204}$$

and

$$P_-(\boldsymbol{p}) = -\sum_{\alpha=1}^{2} v_{\boldsymbol{p}}^{(\alpha)} \otimes \bar{v}_{\boldsymbol{p}}^{(\alpha)} = \frac{\not{p}-m}{\sqrt{2m(m+\omega_{\boldsymbol{p}})}} \frac{1-\gamma^0}{2} \frac{\not{p}-m}{\sqrt{2m(m+\omega_{\boldsymbol{p}})}} = \frac{m-\not{p}}{2m}, \tag{205}$$

respectively either by the help of the definition (147) or simply by using eqs. (199).

The four linearly independent plane wave solutions of the free Dirac equation corresponding to a given (canonical) energy-momentum p^μ represent the two spin polarization states of a fermion of spin half and its anti-particle. The separation of the particle and the anti-particle modes is achieved in the spinor space as in the case of the first order formalism of the Klein-Gordon equation. The momentum-dependence of the projectors (204)-(205) indicates that the particle and anti-particles axes in the spinor space are momentum-dependent, too.

The spinors of the standard and the chiral representation are related by $\psi = U^\dagger\psi_{ch}$. It is advantageous to use the parametrization $\psi = (\phi_+, \phi_-)$ in standard representation because the combinations $\phi_\pm = \eta \pm \xi$ are such that ϕ_+ (ϕ_-) is made up dominantly by particle (anti-particle) modes for non-relativistic particles, $|\boldsymbol{p}| \ll m$, according to (200). The space inversion which preserves the piece $i\boldsymbol{\gamma}\boldsymbol{\nabla}$ of the Dirac equation should not change the eigenvalue of γ^0 neither in order to keep the covariance of the equation. Furthermore, time inversion and charge conjugation change the sign of ∂_0 and should therefore flip the eigenvalue of γ^0.

2. Spin density matrix

The state of the spin is represented by the spin density matrix, usually written in the non-relativistic case as

$$\rho = \frac{1}{2}(\mathbb{1} + \boldsymbol{a}\boldsymbol{\sigma}), \tag{206}$$

giving the expectation value

$$\langle \boldsymbol{S} \rangle = \mathrm{tr}[\rho \boldsymbol{S}] = \frac{\boldsymbol{a}}{2}, \tag{207}$$

and the state is mixed if

$$\mathrm{tr}\rho^2 = \frac{1}{4}\mathrm{tr}(\mathbb{1} + \boldsymbol{a}^2 + 2\boldsymbol{a}\boldsymbol{\sigma}) = \frac{1+\boldsymbol{a}^2}{2} < \mathrm{tr}\rho = 1, \tag{208}$$

i.e., $\boldsymbol{a}^2 < 1$. The relativistically covariant generalization for the particle state (202), is the density matrix

$$\rho = u^{(\alpha)}_{\boldsymbol{p}} \otimes \bar{u}^{(\alpha)}_{\boldsymbol{p}}. \tag{209}$$

The expectation value of the spin operator, given by eq. (197),

$$\langle \boldsymbol{S} \rangle = \int d^3x \psi^*(\boldsymbol{x}) \boldsymbol{S} \psi(\boldsymbol{x}) = \int d^3x \bar{\psi}(\boldsymbol{x}) \gamma^0 \boldsymbol{S} \psi(\boldsymbol{x}), \tag{210}$$

leads to the relativistically covariant expression

$$\langle \boldsymbol{S} \rangle = \mathrm{tr}[\rho \gamma^0 \boldsymbol{S}]. \tag{211}$$

It is sometime useful to express this expectation value in a different form whose derivation starts with an expression of γ^5 involving the Levi-Civita tensor,

$$\gamma^5 = i\gamma^0\gamma^1\gamma^2\gamma^3 = -\frac{i}{4!}\epsilon_{\mu\nu\rho\sigma}\gamma^\mu\gamma^\nu\gamma^\rho\gamma^\sigma, \tag{212}$$

where the minus sign arises from $\epsilon^{0123} = 1$, $\epsilon_{\mu\nu\rho\sigma} = -\epsilon^{\mu\nu\rho\sigma}$. This equation, multiplied by two Dirac matrices,

$$\gamma^5\gamma^\kappa\gamma^\lambda = -\frac{i}{4!}\epsilon_{\mu\nu\rho\sigma}\gamma^\mu\gamma^\nu\gamma^\rho\gamma^\sigma\gamma^\kappa\gamma^\lambda, \tag{213}$$

and evaluated by permuting the indices μ' and ν' to the rightmost poisition in the Levi-Civita symbol results in

$$\gamma^5\gamma_\kappa\gamma_\lambda = -\frac{i}{2}\epsilon_{\mu\nu\lambda\kappa}\gamma^\mu\gamma^\nu, \tag{214}$$

according to eq. (147). This result is now used in the third step of the chain of equations

$$\frac{i}{2}\epsilon^{\mu\nu\rho\sigma}\sigma_{\rho\sigma}=\frac{i}{2}\epsilon^{\mu\nu\rho\sigma}\frac{i}{2}[\gamma_\rho,\gamma_\sigma]=-\frac{1}{2}\epsilon^{\mu\nu\rho\sigma}\gamma_\rho\gamma_\sigma=\gamma^0\gamma^1\gamma^2\gamma^3\gamma^\mu\gamma^\nu=\gamma^5\sigma^{\mu\nu}, \tag{215}$$

allowing us to bring the spin operator into the form

$$S^j=\frac{1}{2}\epsilon^{jk\ell}\sigma^{k\ell}=\frac{1}{2}\epsilon^{0j\rho\sigma}\sigma_{\rho\sigma}=-i\frac{1}{2}\gamma^5\sigma^{0j}=\gamma^5\gamma^0\gamma^j. \tag{216}$$

Finally the anti-commutator, $\{\gamma^5,\gamma^\mu\}=0$, yields an alternative form

$$\langle\boldsymbol{S}\rangle=-\mathrm{tr}[\rho\gamma^5\boldsymbol{\gamma}] \tag{217}$$

for the spin expectation value.

To find the relativistic generalization of the density matrix (206) we introduce a pseudo-vector, $a^\mu=(a^0,\boldsymbol{a})$ which remains invariant under space inversion and becomes $a_0^\mu=(0,2\langle\boldsymbol{S}\rangle)$ in the rest frame, $p_0^\mu=(m,\boldsymbol{0})$. Such a structure yields the orthogonality, $ap=0$, and invariant length square $a^2=-2\langle\boldsymbol{S}\rangle^2$. Though the explicit form of a^μ is not needed below one can easily find it by performing the Lorentz boost,

$$a^0=\frac{a_0^0-va_{0\|}}{\sqrt{1-v^2}},\quad a_\|=\frac{a_{0\|}-va_0^0}{\sqrt{1-v^2}}, \tag{218}$$

with velocity $\boldsymbol{v}=-\boldsymbol{p}/p^0$ where $a_{0\|}=\boldsymbol{a}_0\boldsymbol{p}/|\boldsymbol{p}|$. The result of the boost is

$$a^0=2\frac{|\boldsymbol{p}|}{m}\langle S_\|\rangle,\quad \boldsymbol{a}_\perp=2\langle\boldsymbol{S}_\perp\rangle,\quad a_\|=2\frac{p^0}{m}\langle S_\|\rangle \tag{219}$$

where

$$S_\|=\frac{\boldsymbol{S}\boldsymbol{p}}{|\boldsymbol{p}|},\quad S_\perp=\left(\mathbb{1}-\frac{\boldsymbol{p}\otimes\boldsymbol{p}}{\boldsymbol{p}^2}\right)\boldsymbol{S}. \tag{220}$$

The simplest example is the density matrix of an non-polarized state, $\langle\boldsymbol{a}\rangle=0$ of a particle. It should be a linear expression of the four momentum, p^μ. The form of the projector operator (204) suggests the form

$$\rho=\frac{\not{p}+m}{4m}. \tag{221}$$

The equations

$$\mathrm{tr}[\gamma^{\mu_1}\cdots\gamma^{\mu_{2n+1}}]=0,\quad \mathrm{tr}[\gamma^\mu\gamma^\nu]=\frac{1}{2}\mathrm{tr}[\{\gamma^\mu\gamma^\nu\}]=4g^{\mu\nu} \tag{222}$$

can be used to check $\langle\boldsymbol{S}\rangle=0$. The density matrix of a partially polarized state is more involved, it will be sought in the form,

$$\rho=\frac{1}{8m^2}(\not{p}+m)\rho_0(a)(\not{p}+m) \tag{223}$$

where $\rho_0(0) = \mathbb{1}$. By assuming that $\rho_0(a)$ is a linear function of a^μ we write

$$\rho_0(a) = \mathbb{1} + z\gamma^5 \not{a} \tag{224}$$

cf. (217) and determine the scalar coefficient z by evaluating the spin average in the rest frame where $a_0^\mu = (0, \boldsymbol{a}_0)$ and

$$\rho = \frac{1}{8}(\mathbb{1} + \gamma^0)(\mathbb{1} - z\gamma^5 \boldsymbol{a}\boldsymbol{\gamma}_0)(\mathbb{1} + \gamma^0) = \frac{1}{4}(\mathbb{1} + \gamma^0)(\mathbb{1} - z\gamma^5 \not{a}_0). \tag{225}$$

The corresponding spin expectation value,

$$\langle \boldsymbol{S} \rangle = -\mathrm{tr}[\rho\gamma^5 \boldsymbol{\gamma}] = -\frac{1}{4}\mathrm{tr}[(\mathbb{1} + \gamma^0)(\mathbb{1} - z\gamma^5 \not{a})\gamma^5 \boldsymbol{\gamma}], \tag{226}$$

can be brought into the form

$$\langle \boldsymbol{S} \rangle = \frac{z}{4}\mathrm{tr}[(\mathbb{1} + \gamma^0)\gamma^5 \not{a}\gamma^5 \boldsymbol{\gamma}] = -\frac{z}{4}\mathrm{tr}[(\mathbb{1} + \gamma^0)\not{a}\boldsymbol{\gamma}] = -\frac{z}{4}\mathrm{tr}[\not{a}\boldsymbol{\gamma}] = z\boldsymbol{a} \tag{227}$$

by the help of eqs. (222). The comparison if this result with eq. (207) yields $z = \frac{1}{2}$,

$$\rho = \frac{1}{8m^2}(\not{p} + m)\left(\mathbb{1} + \frac{1}{2}\gamma^5 \not{a}\right)(\not{p} + m), \tag{228}$$

and the identity $\not{a}\not{p} = ab - \not{p}\not{a} = -\not{p}\not{a}$ leads to the form

$$\rho = \frac{1}{4m}(\not{p} + m)\left(\mathbb{1} + \frac{1}{2}\gamma^5 \not{a}\right) \tag{229}$$

of the density matrix. By starting with a given density matrix of this form the parameter a^μ can be extracted by the equation

$$a^\mu = -2\mathrm{tr}[\rho\gamma^5\gamma^\mu]. \tag{230}$$

To prove this result we start with the identity,

$$\mathrm{tr}\left[(\not{p} + m)\left(\mathbb{1} + \frac{1}{2}\gamma^5 \not{a}\right)\gamma^5\gamma^\mu\right] = \mathrm{tr}\left[\left(\not{p} + \frac{m}{2}\gamma^5 \not{a}\right)\gamma^5\gamma^\mu\right]. \tag{231}$$

To simplify it we use the equation $\mathrm{tr}[\gamma^5\gamma^\mu\gamma^\nu] = 0$ which can easily be checked separately for the cases $\mu = \nu$ and $\mu \neq \nu$ and leads to

$$\mathrm{tr}\left[(\not{p} + m)\left(\mathbb{1} + \frac{1}{2}\gamma^5 \not{a}\right)\gamma^5\gamma^\mu\right] = -\frac{m}{2}\mathrm{tr}[\not{a}\gamma^\mu] = -2ma^\mu \tag{232}$$

Q.E.D. As a cross check we calculate the density matrix in the rest-frame and int the standard representation,

$$\rho \;=\; \frac{1}{4}(\mathbb{1} + \gamma^0)\left(\mathbb{1} + \frac{1}{2}\gamma^5 \not{a}\right)$$

$$= \frac{1}{4}\begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}\left[\mathbb{1} - \frac{1}{2}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 0 & \boldsymbol{a\sigma} \\ -\boldsymbol{a\sigma} & 0 \end{pmatrix}\right] = \frac{1}{4}\begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}\left[\mathbb{1} - \frac{1}{2}\begin{pmatrix} -\boldsymbol{a\sigma} & 0 \\ 0 & \boldsymbol{a\sigma} \end{pmatrix}\right]$$
$$= \frac{1}{2}\begin{pmatrix} \mathbb{1} + \boldsymbol{a\sigma} & 0 \\ 0 & 0 \end{pmatrix} \tag{233}$$

Finally we add that the density matrix for an anti-particle is given by

$$\rho^{(-)} = U_C \rho^{(+)} U_C^*. \tag{234}$$

3. Helicity, chirality, Weyl and Majorana fermions

A Dirac fermion is described by a four component, complex bi-spinor and there are different ways of splitting this eight dimensional vector into two four dimensional ones. We discuss three of them below.

The **helicity** is defined as the projection of the total angular momentum of a particle on its momentum,

$$h_{\boldsymbol{p}} = \frac{\boldsymbol{Jp}}{|\boldsymbol{p}|}. \tag{235}$$

The orbital angular momentum drops from this expression,

$$h_{\boldsymbol{p}} = \frac{\boldsymbol{Sp}}{|\boldsymbol{p}|}, \tag{236}$$

owing to $\boldsymbol{J} = \boldsymbol{L} + \boldsymbol{s}$ and $\boldsymbol{Lp} = 0$, hence the helicity of a Dirac fermion is $\pm\hbar/2$. The helicity commutes with the Hamiltonian (142), $[\boldsymbol{\alpha p} + \beta m, \boldsymbol{Sp}] = 0$, and is thereby conserved, c.f. Table II. It is furthermore invariant under spatial rotations and offers a natural basis to handle the spin orientations of a free particle. However it changes under Lorentz boosts which mixes the helicity component. This can be anticipated by noting that it changes sign when we perform a Lorentz boost with a velocity which is parallel to and larger than the particle velocity. The spin projection on the direction of the velocity arises from the components of the tensor $\sigma_{k\ell}$ which are orthogonal to the velocity and thus remain unchanged. But the velocity of the particle changes sign during the boost, together with the helicity. Thus a massive fermion can not be composed exclusively from a given helicity components.

The spinor representation of the special Lorentz group in which a massless particle is found can be equipped with a quantum number, called chirality. It is defined as the eigenvalue of the Dirac matrix γ^5 according to the form (188) in chiral representation. Thus

$$R = \frac{1}{2}(\mathbb{1} + \gamma^5), \quad L = \frac{1}{2}(\mathbb{1} - \gamma^5), \tag{237}$$

TABLE II: Invariance properties of the splitting of the bi-spinor space.

Invariance	Helicity	Chirality	Majorana fermion
$L_+^\uparrow$	$\times$	✓	✓
Time evolution of a free particle	✓	$m \neq 0:\ \times$; $m = 0$: ✓	✓

project onto **chirality** $+1$ and -1, called right-handed and left-handed states, respectively. Chirality is a Lorentz invariant quantity but is conserved by massless particles only.

Both the helicity and the chirality are conserved by masless particles hence one wonders whether there is a relation between these quantities. The equation of motion for massless particles,

$$(\partial_0 - \boldsymbol{\nabla}\boldsymbol{\sigma})\eta = 0, \quad (\partial_0 + \boldsymbol{\nabla}\boldsymbol{\sigma})\xi = 0, \tag{238}$$

cf. eqs. (183), written in the form

$$i\partial_0\eta = -\boldsymbol{p}\boldsymbol{\sigma}\eta, \quad i\partial_0\xi = \boldsymbol{p}\boldsymbol{\sigma}\xi, \tag{239}$$

shows that the helicity and the chirality are identical for positive energy massless excitations. The name helicity refers to the opposite direction of the rotation of the helicity ($\boldsymbol{S} = \boldsymbol{\sigma}/2$) for the left and the right handed spinors.

Weyl fermions, the left or right handed spinors of the special Lorentz group, $\xi = R\psi$ and $\eta = L\psi$, are massless fermions which respect no space inversion. They can be described either by a spinor ξ or η exclusively and the corresponding equation of motions, (238), can be derived from the Lagrangian

$$L_\eta = \eta^\dagger(i\partial_0 - i\boldsymbol{\nabla}\boldsymbol{\sigma})\eta, \quad L_\xi = \xi^\dagger(i\partial_0 + i\boldsymbol{\nabla}\boldsymbol{\sigma})\xi. \tag{240}$$

The Weyl fermions are massless and the first equation in (238) was sometime called neutrino equation, by now a historic name from the time when neutrinos were thought to be massless.

One can find a basis transformation

$$U_M = \frac{1}{2}\begin{pmatrix} 1+\sigma_2 & i(\sigma_2 - 1) \\ i(1-\sigma_2) & 1+\sigma_2 \end{pmatrix} \tag{241}$$

which brings the gamma matrices into the imaginary form, $\gamma_M^\mu = U\gamma_{ch}^\mu U^\dagger$, in particular

$$\gamma_M^0 = \begin{pmatrix} 0 & \sigma_2 \\ \sigma_2 & 0 \end{pmatrix}, \quad \gamma_M^1 = \begin{pmatrix} i\sigma_1 & 0 \\ 0 & i\sigma_1 \end{pmatrix}, \quad \gamma_M^2 = \begin{pmatrix} 0 & \sigma_2 \\ -\sigma_1 & 0 \end{pmatrix}, \quad \gamma_M^3 = \begin{pmatrix} i\sigma_3 & 0 \\ 0 & i\sigma_3 \end{pmatrix}, \tag{242}$$

and thereby renders the Dirac equation real. Its solution preserves the phase and represents a **Majorana fermion**.

One may wonder if the four real functions of a Majorana fermion could be compressed in a Weyl fermion. The difficulty is that the realness of the Dirac fermion wave function, $\psi^* = \psi$, is not a Lorentz invariant condition. But the bi-spinor

$$\psi_M = \begin{pmatrix} \xi \\ -i\sigma_2\xi^* \end{pmatrix}, \tag{243}$$

given in the chiral representation is brought into a real bi-spinor by the basis transformation (241). While a Weyl fermion can not have mass the Majorana fermion which has the same number of real components can. In fact, the Lagrangian

$$L_M = \frac{1}{2}\bar{\psi}_M(i\partial\!\!\!/ - m)\psi_M, \tag{244}$$

is manifest Lorentz invariant and becomes

$$L_M = \frac{1}{2}[\xi^\dagger i(\partial_0 - \boldsymbol{\nabla\sigma})\xi - m\xi^{\rm tr} i\sigma_2\xi] + c.c. \tag{245}$$

in the chiral representation becomes where c.c. stands for complex conjugation. The corresponding equation of motion is massive,

$$(\partial_0 - \boldsymbol{\nabla\sigma})\xi - m\sigma_2\xi^* = 0. \tag{246}$$

The precision tests of the Standard Model should soon reveal if the neutrino masses are generated for Weyl or Majorana neutrinos.

The construction of the Majorana fermions shows the highly non-trivial structure of the Dirac matrices, arising from the identity (166). One may go even further and use eq. (168) to represent the space-time coordinate vectors by spinors, $x^\mu \leftrightarrow X^{a\dot{a}}(x) \leftrightarrow \nu^a\tilde{\nu}^{\dot{a}}$. To have four dimensional objects one sets $\tilde{\nu} = \nu^*$ and parametrizes the coordinate in terms of the twistor ν,

$$X^{a\dot{a}}(x) = \nu^a\nu^{*\dot{a}}. \tag{247}$$

However this reduces the dimensionality of X to three because X remains invariant under the global phase change, $\nu \to e^{i\alpha}\nu$. This is in agreement with the observation that the rank of $\nu^a\nu^{*\dot{a}}$ is one, giving $\det X = x^\mu x_\mu = 0$, the twistor parametrization is available for light-like vectors only.

D. Non-relativistic limit

Let us find the non-relativistic limit of the Dirac equation with a vector potential, written as

$$\frac{i}{c}\partial_t\psi = \left[\boldsymbol{\alpha}(-i\boldsymbol{\nabla} - \frac{e}{c}\boldsymbol{A}) + \beta mc + \frac{e}{c}A_0\right]\psi, \tag{248}$$

where the units $c \neq 1$ are reintroduced. We shall work in the standard representation, (149), and write the equation for the bi-spinor $\psi = (\phi, \chi)$,

$$\begin{aligned} i\partial_t\phi &= c\boldsymbol{\sigma\pi}\chi + (eA_0 + mc^2)\phi, \\ i\partial_t\chi &= c\boldsymbol{\sigma\pi}\phi + (eA_0 - mc^2)\chi, \end{aligned} \tag{249}$$

where the momentum $\pi = \boldsymbol{p} - \frac{e}{c}\boldsymbol{A}$ is used. To find regular time dependence in the non-relativistic limit for an electron we separate the rest mass energy in the solution and use $\Phi = e^{imc^2t}\phi$ and $X = e^{imc^2t}\chi$ in the equation of motion

$$\begin{aligned} i\partial_t\Phi &= c\boldsymbol{\sigma\pi}X + eA_0\Phi, \\ i\partial_t X &= c\boldsymbol{\sigma\pi}\Phi + (eA_0 - 2mc^2)X. \end{aligned} \tag{250}$$

As long as we can neglect the time derivative and eA_0 in the second equation compared to mc^2 we can use the approximation

$$X = \frac{\boldsymbol{\sigma\pi}}{2mc}\Phi \tag{251}$$

for the small component of the bi-spinor. Inserting this relation into the first equation we arrive at Pauli equation for the large component,

$$i\partial_t\Phi = \left[\frac{(\boldsymbol{\sigma\pi})^2}{2m} + eA_0\right]\Phi, \tag{252}$$

suggesting that the influence of the small component is the replacement $\boldsymbol{p} \to \boldsymbol{\sigma\pi}$ in the Schrödinger equation. The identity (163) gives

$$(\boldsymbol{\sigma\pi})^2 = \sigma_j\sigma_k\pi_j\pi_k = \boldsymbol{\pi}^2 - \frac{e}{c}\boldsymbol{B} \tag{253}$$

with $\boldsymbol{B} = \boldsymbol{\nabla} \times \boldsymbol{A}$ and the Pauli equation,

$$i\partial_t\Phi = \left[\frac{(\boldsymbol{p} - \frac{e}{c}\boldsymbol{A})^2}{2m} - \frac{e}{2mc}\boldsymbol{\sigma B} + eA_0\right]\Phi. \tag{254}$$

The first term appears in the Schrödinger equation and it reads for a homogeneous magnetic field, $\boldsymbol{A}(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x} \times \boldsymbol{B}$, as

$$H_{Sch} = \frac{\boldsymbol{p}^2}{2m} + \frac{e\boldsymbol{Ap}}{mc} = \frac{\boldsymbol{p}^2}{2m} - \frac{e\boldsymbol{LB}}{2mc}. \tag{255}$$

This form reveals an angular momentum-magnetic field interaction with energy

$$H_L = -\mu_B \frac{\boldsymbol{LB}}{\hbar}, \tag{256}$$

where $\mu_b = e\hbar/2mc$ is the Bohr magneton. The second term in the equation of motion, (254), can be written as

$$H_S = -g_S \mu_B \frac{\boldsymbol{SB}}{\hbar} \tag{257}$$

where $g_S = 2$ is the gyromagnetic factor. The lesson of the Pauli equation is that the magnetic field can distinguish the orbital and the spin contribution to the total angular momentum, $\boldsymbol{J} = \boldsymbol{L} + \boldsymbol{S}$, because the angular momentum-magnetic field coupling is

$$H_J = -\mu_B \frac{(\boldsymbol{L} + 2\boldsymbol{S})\boldsymbol{B}}{\hbar} \tag{258}$$

in the non-relativistic limit. An electron continuously emits and absorbs photons during its propagation and these processes change its angular momentum and generate a small non-trivial value for $g_S - 2$. This value in known experimentally up to twelve decimals and has been checked by $\mathcal{O}\left(\alpha^4\right)$ perturbation calculation. The agreement represents one of the precision tests of quantum electrodynamics.

Such a treatment of the non-relativistic limit is flowed. One problem is the issue of the velocity: The Heisenberg equation for the coordinate operator,

$$\frac{i}{c}\partial_t \boldsymbol{x} = [\boldsymbol{x}, \boldsymbol{\alpha p} + \beta mc^2] = i\boldsymbol{\alpha}, \tag{259}$$

shows that the particle moves with the speed of light because the spectrum of α_j is ± 1. The average speed becomes small in the non-relativistic limit due to the frequent flip of sign of the eigenvalues of α_j as the state oscillates between the different eigenstates of β. This is actually the context how Zitterbewegung has been identified as the result of the interference between positive and negative energy solutions of the Dirac equation.

This result is reasonable, argued Dirac, since the velocity is obtained by two successive measurement in time of the coordinate. The Heisenberg uncertainty principle introduces a large spread for the momentum when the two coordinate measurements are carried out in short time difference. The practically uniformly distributed momentum is dominated by large values of momentum, corresponding to the velocity of light. There must obviously a way to get through this complication and arrive at the usual description of a non-relativistic electron in Schrödinger's wave mechanics.

Another problem of the derivation of the Pauli equation is that the time derivative of the small component, neglected in the derivation, would introduce higher order time derivatives in the

equation of motion. Such higher order derivative terms, whatever weak they are in the equation of motion, produce important effects in sufficiently long time. This can be seen by recalling Ostrogadsky's theorem, stating that the energy of a classical system whose equation of motion is canonical, i.e. can be derived from a Lagrangian or Hamiltonian, is unbounded from below. Thus we loose the stability of the classical equations of motion higher than second order. Since we do not know the fundamental equations in physics, all we have are effective equations, valid in certain resolution window. These equations are always non-local in time. The non-local equations represent a far too hard problem to solve hence one usually applies the gradient expansion, truncated at a certain order. The lesson of Ostrogadky's theorem is that this procedure is acceptable in dealing with non-localities in space but can not be applied in time.

A more physical approach of the non-relativistic limit is offered by the Foldy-Wouthuysen transformation which is a basis transformation, $|\psi\rangle \to S|\psi\rangle$ to decouple the small and the large components. We look for the the similarity transformation of the form

$$S = e^{-\frac{\boldsymbol{\gamma p}}{|\boldsymbol{p}|}\theta} = \mathbb{1}\cos\theta + \frac{\boldsymbol{\gamma p}}{|\boldsymbol{p}|}\sin\theta, \tag{260}$$

where the sign of the sinus is positive owing to $(\boldsymbol{\gamma p})^2 = -\boldsymbol{p}^2$. It transforms the Hamiltonian into

$$\begin{aligned} S(\boldsymbol{\alpha p} + \beta m)S^{-1} &= \left(\mathbb{1}\cos\theta + \frac{\boldsymbol{\gamma p}}{|\boldsymbol{p}|}\sin\theta\right)(\boldsymbol{\alpha p} + \beta m)\left(\mathbb{1}\cos\theta - \frac{\boldsymbol{\gamma p}}{|\boldsymbol{p}|}\sin\theta\right) \\ &= (\boldsymbol{\alpha p} + \beta m)\left(\mathbb{1}\cos\theta - \frac{\boldsymbol{\gamma p}}{|\boldsymbol{p}|}\sin\theta\right)^2 \\ &= (\boldsymbol{\alpha p} + \beta m)\left(\mathbb{1}\cos 2\theta - \frac{\boldsymbol{\gamma p}}{|\boldsymbol{p}|}\sin 2\theta\right) \\ &= \boldsymbol{\alpha p}\left(\mathbb{1}\cos 2\theta - \frac{m}{|\boldsymbol{p}|}\sin 2\theta\right) + \beta(m\cos 2\theta + |\boldsymbol{p}|\sin 2\theta), \end{aligned} \tag{261}$$

where the equation $p^j p^k\{\gamma^j, \alpha^k\} = p^j p^k \gamma^0[\gamma^j, \gamma^k] = 0$ was used in changing the order of the frist two factors and in arriving at the second line. The choice

$$\sin 2\theta = \frac{|\boldsymbol{p}|}{\omega_p}, \quad \cos 2\theta = \frac{m}{\omega_p}, \quad \tan 2\theta = \frac{|\boldsymbol{p}|}{m}, \tag{262}$$

decouples the positive and negative energy modes since

$$H_{FW} = \beta\omega_p. \tag{263}$$

The form of the projectors eqs. (204)-(205) in this base,

$$\begin{aligned} SP_\pm(\boldsymbol{p})S^{-1} &= \left(\mathbb{1}\cos\theta + \frac{\boldsymbol{\gamma p}}{|\boldsymbol{p}|}\sin\theta\right)\frac{m \pm (\beta\omega_p - \boldsymbol{\gamma p})}{2m}\left(\mathbb{1}\cos\theta - \frac{\boldsymbol{\gamma p}}{|\boldsymbol{p}|}\sin\theta\right) \\ &= \frac{m \mp \boldsymbol{\gamma p}}{2m} \pm \frac{\beta\omega_p}{2m}\left(\mathbb{1}\cos\theta - \frac{\boldsymbol{\gamma p}}{|\boldsymbol{p}|}\sin\theta\right)^2 \end{aligned}$$

$$
\begin{aligned}
&= \frac{m \mp \boldsymbol{\gamma p}}{2m} \pm \frac{\beta\omega_p}{2m}\left(\mathbb{1}\frac{m}{\omega_p} - \frac{\boldsymbol{\gamma p}}{|\boldsymbol{p}|}\frac{|\boldsymbol{p}|}{\omega_p}\right) \\
&= \frac{\mathbb{1} \pm \beta}{2},
\end{aligned} \tag{264}
$$

shows explicitly that the positive and negative energy plane wave belong to exactly the upper or lower components of the bi-spinor in this representation.

The velocity operator in this basis, given by the Heisenberg equation of for the coordinate,

$$
\begin{aligned}
\frac{1}{c}\partial_t \boldsymbol{x} &= -i[\boldsymbol{x}, \beta\omega_p] \\
&= \beta \boldsymbol{\nabla}_p \omega_p
\end{aligned} \tag{265}
$$

is the group velocity up to the sign of the energy, the eigenvalue of β.

The momentum dependence of the basis transformation indicates that the coordinate operator of the Foldy-Wouthuysen representation belongs to

$$
\begin{aligned}
\boldsymbol{x}_{FW} &= S\boldsymbol{x}S^{-1} \\
&= \left(\mathbb{1}\cos\theta + \frac{\boldsymbol{\gamma p}}{|\boldsymbol{p}|}\sin\theta\right)\boldsymbol{x}\left(\mathbb{1}\cos\theta - \frac{\boldsymbol{\gamma p}}{|\boldsymbol{p}|}\sin\theta\right) \\
&= \boldsymbol{x} + i\boldsymbol{a}
\end{aligned} \tag{266}
$$

in the original Dirac basis with

$$
\boldsymbol{a} = -\boldsymbol{\nabla}_p\left(\frac{\boldsymbol{\gamma p}}{|\boldsymbol{p}|}\sin\theta\right). \tag{267}
$$

E. External field

The mixing of the particle-anti particle modes are considered below in the presence of an external potential.

1. Klein paradox

The first example to consider is the one-dimensional motion in the presence of a potential barrier, $U(z) = U_0\Theta(z)$, shown in Fig. 3. The wave function,

$$
\psi(t, z) = \chi(z)e^{-itE}, \tag{268}
$$

satisfies the Dirac equation,

$$
[\gamma^0(E - U(z)) + i\gamma^z\nabla_z - m]\chi(z) = 0. \tag{269}
$$

We assume that $U_0 > 2m$ and $m < E < -m + U_0$ to assure the extended incoming particle and transmitted anti-particle modes in the stationary state. The wave function for $z < 0$ is the sum of the incident and the reflected waves, $\chi = \chi_i + \chi_r$, with

$$\begin{aligned}
\chi_i(z) &= e^{ipz} \begin{pmatrix} 1 \\ 0 \\ \frac{p}{m+E} \\ 0 \end{pmatrix} \\
\chi_r(z) &= be^{-ipz} \begin{pmatrix} 1 \\ 0 \\ -\frac{p}{m+E} \\ 0 \end{pmatrix} + b'e^{-ipz} \begin{pmatrix} 0 \\ 1 \\ 0 \\ \frac{p}{m+E} \end{pmatrix},
\end{aligned} \tag{270}$$

according to eq. (202). The solution for $z > 0$ describes the transmitted wave,

$$\chi_t(z) = de^{ip'z} \begin{pmatrix} 1 \\ 0 \\ \frac{p'}{m+E-U_0} \\ 0 \end{pmatrix} + d'e^{ip'z} \begin{pmatrix} 0 \\ 1 \\ 0 \\ -\frac{p'}{m+E-U_0} \end{pmatrix}. \tag{271}$$

The momenta are defined by the equations $E^2 = m^2 + p^2$ and $(E - U_0)^2 = m^2 + p'^2$. The potential is spin independent and the spin flip is excluded, $b' = d' = 0$.

To find the matching condition at the singularity of the potential we write the Dirac equation as

$$\nabla_z \chi(z) = i\gamma^z [m - \gamma^0 (E - U(z))]\chi(z) \tag{272}$$

and integrate within the interval $-\epsilon < z < \epsilon$. The result is that the discontinuity of the wave function is given by

$$\mathrm{Disc}\chi(0) = i\gamma^z \lim_{\epsilon \to 0} \int_{-\epsilon}^{\epsilon} [m - \gamma^0 (E - U(z))]\chi(z). \tag{273}$$

The right hand side of vanishing for bounded potential therefore we have the matching condition is

$$1 + b = d \tag{274}$$

for the first component of the spinor and

$$(1 - b)\frac{p}{m + E} = d\frac{p'}{m + E - U_0} \tag{275}$$

for the third component. We write the latter as

$$1 - b = d\xi \tag{276}$$

with

$$\xi = \frac{p'}{p}\frac{m+E}{m+E-U_0}, \tag{277}$$

yielding

$$b = \frac{1-\xi}{1+\xi}, \quad d = \frac{2}{1+\xi}. \tag{278}$$

The reflection and transmission coefficients are defined by means of the current,

$$j^z = \bar{\psi}\gamma^z\psi = \psi^\dagger\gamma^0\gamma^z\psi = \chi^\dagger \begin{pmatrix} 0 & \sigma^z \\ \sigma^z & 0 \end{pmatrix} \chi, \tag{279}$$

as

$$R = -\frac{j_r}{j_i}, \quad T = \frac{j_t}{j_i}. \tag{280}$$

Since

$$j_i = 2\frac{p}{m+E}, \quad j_r = -2|b|^2\frac{p}{m+E}, \quad j_t = 2|d|^2\frac{p'}{m+E-U_0}, \tag{281}$$

we have

$$R = |b|^2, \quad T = |d|^2\frac{p'}{p}\frac{m+E-U_0}{m+E}. \tag{282}$$

Note that the current is conserved,

$$j_i(0) + j_r(0) = \frac{2(1-|b|^2)p}{m+E} = \frac{2(1-|b|^2)p'}{\xi(m+E-U_0)} = j_t(0)\frac{1-|b|^2}{\xi|d|^2} = j_t(0), \tag{283}$$

and $R + T = 1$. Since $\xi < 0$ in the considered energy interval we have $R > 1$ and $T < 0$, known as the Klein paradox. Such a process is absent in the second quantized formalism of quantum field theories.

2. Spherical potential

The solution Dirac equation with a time-independent potential,

$$i\partial_0\psi(x) = [\boldsymbol{\alpha}\boldsymbol{p} + \beta m + U(\boldsymbol{x})]\psi(x), \tag{284}$$

is sought by factorizing an oscillatory time dependence by the help of the ansatz

$$\psi(t,\boldsymbol{x}) = \psi(\boldsymbol{x})e^{-iEt}. \tag{285}$$

The space-dependent wave function satisfies the stationary Dirac equation,

$$[\boldsymbol{\alpha p} + \beta m + U(\boldsymbol{x})]\psi(\boldsymbol{x}) = E\psi(\boldsymbol{x}). \tag{286}$$

In case of a spherically symmetric potential, $U = U(r)$, one obviously uses polar coordinate system and the directional dependence of the wave function is provided by using the spinor spherical harmonics

$$\mathcal{Y}^{\ell}_{J,M} = \sum_{m+\sigma=M} (\ell, m, \frac{1}{2}, \sigma | J, M) Y^{\ell}_{m}(\theta,\phi)\phi(\sigma), \tag{287}$$

where the usual spherical harmonics satisfy the equgenvalue conditions

$$\boldsymbol{L}^2 Y^{\ell}_{m} = \ell(\ell+1)Y^{\ell}_{m}, \quad L_3 Y^{\ell}_{m} = mY^{\ell}_{m}, \tag{288}$$

and the spin states are defined by the conditions

$$\boldsymbol{S}^2\phi(\sigma) = \frac{3}{4}\phi(\sigma), \quad S_3\phi(\sigma) = \sigma\phi(\sigma). \tag{289}$$

The rule of addition of angular momentum, $\ell \otimes \frac{1}{2} = \ell - \frac{1}{2} \oplus \ell + \frac{1}{2}$, shows that a given total angular momentum state, characterized by a given J, M quantum number pair corresponds to a two-dimensional subspace with orbital angular momentum $\ell = J \pm \frac{1}{2}$. Owing to the transformation rules under space inversion,

$$PY^{\ell}_{m}(\theta,\phi) = (-1)^{\ell} Y^{\ell}_{m}(\theta,\phi), \quad P\boldsymbol{S}P^{-1} = \boldsymbol{S} \tag{290}$$

these two states with different orbital angular momentum have different parity,

$$P\mathcal{Y}^{\ell}_{J,M}(\theta,\phi) = (-1)^{\ell}\mathcal{Y}^{\ell}_{J,M}(\theta,\phi). \tag{291}$$

This is an indirect justification of regrouping two irreducible representations of the proper Lorentz group together to find a representation of the space inversion. In particular,

$$P\begin{pmatrix}\phi(t,\boldsymbol{x})\\ \chi(t,\boldsymbol{x})\end{pmatrix} = i\gamma^0\begin{pmatrix}\phi(t,-\boldsymbol{x})\\ \chi(t,-\boldsymbol{x})\end{pmatrix} = i\begin{pmatrix}\phi(t,-\boldsymbol{x})\\ -\chi(t,-\boldsymbol{x})\end{pmatrix} \tag{292}$$

in the standard representation. Hence the space-dependent component of the bi-spinor with definite parity, $\pi = (-1)^{J\pm\frac{1}{2}}$, can be written into the form

$$\psi^{(\pm)}_{J,M}(r,\theta,\phi) = \frac{1}{r}\begin{pmatrix}u_{J\pm\frac{1}{2}}(r)\mathcal{Y}^{J\pm\frac{1}{2}}_{J,M}(\theta,\phi)\\ iv_{J\mp\frac{1}{2}}(r)\mathcal{Y}^{J\mp\frac{1}{2}}_{J,M}(\theta,\phi)\end{pmatrix}. \tag{293}$$

To write the stationary Dirac equation in terms of this bi-spinor we need a more appropriate expression for $\boldsymbol{p\alpha}$. By the help of the radial unit vector, $\boldsymbol{n} = \boldsymbol{x}/r$, one can write

$$\begin{aligned}(\boldsymbol{n\alpha})(\boldsymbol{p\alpha}) &= \begin{pmatrix} 0 & \boldsymbol{n\sigma} \\ \boldsymbol{n\sigma} & 0 \end{pmatrix} \begin{pmatrix} 0 & \boldsymbol{p\sigma} \\ \boldsymbol{p\sigma} & 0 \end{pmatrix} \\ &= [\boldsymbol{np} + i(\boldsymbol{n}\times\boldsymbol{p})\boldsymbol{\sigma}] = \left(\boldsymbol{np} + \frac{i}{r}\boldsymbol{L\sigma}\right) = \left(-i\partial_r + \frac{i}{r}\boldsymbol{L\sigma}\right), \end{aligned} \tag{294}$$

yielding

$$\boldsymbol{p\alpha} = (\boldsymbol{n\alpha})^2(\boldsymbol{p\alpha}) = \boldsymbol{n\alpha}\left(-i\partial_r + \frac{i}{r}\boldsymbol{L\sigma}\right). \tag{295}$$

The next step is to work out the action of $\boldsymbol{n\alpha}$ and $\boldsymbol{L\sigma}$ on the spherical harmonics. We start with the eigenvalue equation for the bi-spinor,

$$\begin{aligned}\boldsymbol{L}^2\psi^{(\pm)}_{J,M} &= \begin{pmatrix} (J\pm\frac{1}{2})(J\pm\frac{1}{2}+1) \\ (J\mp\frac{1}{2})(J\mp\frac{1}{2}+1) \end{pmatrix} \psi^{(\pm)}_{J,M} = \begin{pmatrix} J^2+\frac{1}{4}\pm J + J \pm \frac{1}{2} \\ J^2+\frac{1}{4}\mp J + J \mp \frac{1}{2} \end{pmatrix} \psi^{(\pm)}_{J,M} \\ &= \left[J(J+1) + \frac{1}{4} \pm \beta\left(J+\frac{1}{2}\right)\right]\psi^{(\pm)}_{J,M}, \end{aligned} \tag{296}$$

giving

$$\boldsymbol{L\sigma}\psi^{(\pm)}_{J,M} = (\boldsymbol{J}^2 - \boldsymbol{L}^2 - \boldsymbol{S}^2)\psi^{(\pm)}_{J,M} = \left[-1 \mp \beta\left(J+\frac{1}{2}\right)\right]\psi^{(\pm)}_{J,M}. \tag{297}$$

The action of the matrix $\boldsymbol{n\sigma}$ on $\mathcal{Y}^{J\pm\frac{1}{2}}_{J,M}$ is found in three steps.

1. First we relate in actions of $\boldsymbol{n\sigma}$ and $\boldsymbol{L\sigma}$. The equations

$$\begin{aligned}(\boldsymbol{n\sigma})(\boldsymbol{L\sigma}) &= i(\boldsymbol{n}\times\boldsymbol{L})\boldsymbol{\sigma} = \frac{i}{r}\sigma^a\epsilon^{abc}x^b\epsilon^{cde}x^dp^e = \frac{i}{r}[\boldsymbol{x}(\boldsymbol{xp}) - \boldsymbol{p}\boldsymbol{x}^2]\boldsymbol{\sigma}, \\ (\boldsymbol{L\sigma})(\boldsymbol{n\sigma}) &= \frac{i}{r}\epsilon^{abc}x^bp^c\epsilon^{ade}x^d\sigma^e = \frac{i}{r}(\epsilon^{abc}x^b\epsilon^{ade}x^dp^c - i\epsilon^{abc}x^b\epsilon^{ace})\sigma^e \\ &= \frac{i}{r}[\boldsymbol{x}^2\boldsymbol{p\sigma} - (\boldsymbol{x\sigma})(\boldsymbol{xp}) + 2i\boldsymbol{x\sigma}], \end{aligned} \tag{298}$$

   serve to establish

$$(\boldsymbol{L\sigma})(\boldsymbol{n\sigma}) = -(\boldsymbol{n\sigma})(\boldsymbol{L\sigma}) - 2\boldsymbol{n\sigma}. \tag{299}$$

2. We apply this equation on $\mathcal{Y}^{J\mp\frac{1}{2}}_{J,M}$ and use the lower block of eq. (297) to find

$$(\boldsymbol{L\sigma})(\boldsymbol{n\sigma})\mathcal{Y}^{J\mp\frac{1}{2}}_{J,M} = \left[-1 \mp \left(J+\frac{1}{2}\right)\right]\boldsymbol{n\sigma}\mathcal{Y}^{J\mp\frac{1}{2}}_{J,M}. \tag{300}$$

   The comparison of this result with the upper block of eq. (297),

$$\boldsymbol{L\sigma}\mathcal{Y}^{J\pm\frac{1}{2}}_{J,M} = \left[-1 \mp \left(J+\frac{1}{2}\right)\right]\mathcal{Y}^{J\pm\frac{1}{2}}_{J,M}, \tag{301}$$

shows that $\boldsymbol{n\sigma}\mathcal{Y}_{J,M}^{J\mp\frac{1}{2}}$ and $\mathcal{Y}_{J,M}^{J\pm\frac{1}{2}}$ satisfy the same non-degenerate eigenvalue equations hence

$$\boldsymbol{n\sigma}\mathcal{Y}_{J,M}^{J\mp\frac{1}{2}} = a_\ell \mathcal{Y}_{J,M}^{J\pm\frac{1}{2}}. \tag{302}$$

3. The eigenvalues of $\boldsymbol{n\sigma}$ are ± 1 which requires $|a| = 1$. In the special case $M = J$ this equation reads

$$\boldsymbol{n\sigma}\begin{pmatrix}1\\0\end{pmatrix} Y_{J,J}^{J-\frac{1}{2}} = a_\ell \begin{pmatrix}0\\1\end{pmatrix} Y_{J,J}^{J+\frac{1}{2}}, \quad \boldsymbol{n\sigma}\begin{pmatrix}0\\1\end{pmatrix} Y_{J,J}^{J+\frac{1}{2}} = a_\ell \begin{pmatrix}1\\0\end{pmatrix} Y_{J,J}^{J-\frac{1}{2}} \tag{303}$$

since the spin is parallel and anti-parallel with the orbital momentum on the two sides of the equation. The form $Y_\ell^\ell \sim [-(x+iy)]^\ell$ of the spherical harmonics gives $a_\ell = a = -1$ and

$$\boldsymbol{n\sigma}\mathcal{Y}_{J,M}^{J\mp\frac{1}{2}} = -\mathcal{Y}_{J,M}^{J\pm\frac{1}{2}}. \tag{304}$$

We have now all the necessary elements to work out the Dirac equation in terms of our ansatz. The equations (295) and (297) brings the stationary Dirac equation into the form

$$\left[i\boldsymbol{n\alpha}\left[-\partial_r - \frac{1}{r} \mp \frac{1}{r}\beta\left(J+\frac{1}{2}\right)\right] + \beta m + U\right]\frac{1}{r}\begin{pmatrix} u_{J\pm\frac{1}{2}}\mathcal{Y}_{J,M}^{J\pm\frac{1}{2}} \\ iv_{J\mp\frac{1}{2}}\mathcal{Y}_{J,M}^{J\mp\frac{1}{2}}\end{pmatrix} = E\frac{1}{r}\begin{pmatrix} u_{J\pm\frac{1}{2}}\mathcal{Y}_{J,M}^{J\pm\frac{1}{2}} \\ iv_{J\mp\frac{1}{2}}\mathcal{Y}_{J,M}^{J\mp\frac{1}{2}}\end{pmatrix}. \tag{305}$$

The identity $(\partial_r + \frac{1}{r})\frac{f}{r} = \frac{f'}{r}$ and eq. (304) allow us to further simplify this result,

$$\left[-i\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix}\left[-\partial_r \mp \frac{1}{r}\beta\left(J+\frac{1}{2}\right)\right] + \beta m + U(r)\right]\begin{pmatrix} u_{J\pm\frac{1}{2}} \\ iv_{J\mp\frac{1}{2}}\end{pmatrix} = E\begin{pmatrix} u_{J\pm\frac{1}{2}} \\ iv_{J\mp\frac{1}{2}}\end{pmatrix} \tag{306}$$

and finally arrive at the equation

$$\begin{aligned}\left[-\partial_r \pm \frac{1}{r}\left(J+\frac{1}{2}\right)\right] v_{J\mp\frac{1}{2}} &= (E - m - U)u_{J\pm\frac{1}{2}}, \\ \left[\partial_r \pm \frac{1}{r}\left(J+\frac{1}{2}\right)\right] u_{J\pm\frac{1}{2}} &= (E + m - U)v_{J\mp\frac{1}{2}}.\end{aligned} \tag{307}$$

We can extract an equation for $u_{J\pm\frac{1}{2}}$ by expressing $v_{J\mp\frac{1}{2}}$ from the second equation and inserting it into the first equation,

$$0 = \left\{m + U - E + \left[-\partial_r \pm \frac{1}{r}\left(J+\frac{1}{2}\right)\right]\frac{1}{E+m-U}\left[\partial_r \pm \frac{1}{r}\left(J+\frac{1}{2}\right)\right]\right\} u_{J\pm\frac{1}{2}}. \tag{308}$$

It is instructive to check the free case, $U = 0$,

$$0 = \left[m^2 - E^2 - \partial_r^2 + \frac{1}{r^2}\left(J+\frac{1}{2}\right)\left(J+\frac{1}{2}\pm 1\right)\right] u_{J\pm\frac{1}{2}} \tag{309}$$

which is identical with that Klein-Gordon case.

3. Spin precession

The non-relativistic dynamics of the spin in the presence of a homogeneous magnetic field, $\boldsymbol{B}$, is generated by the modification

$$H = H_0 - \mu_B \boldsymbol{\sigma} \boldsymbol{B} \tag{310}$$

of the spin-less Hamiltonian where $\mu_B = \frac{e\hbar}{2mc}$. The Heisenberg equation for the spin operator, $\boldsymbol{S} = \boldsymbol{\sigma}/2$,

$$i\partial_t \boldsymbol{S} = [\boldsymbol{S}, H], \tag{311}$$

and $[S^j, \sigma^k] = i\epsilon^{jk\ell}\sigma^\ell = 2i\epsilon^{jk\ell}S^\ell$ leads to the equation of motion

$$\partial_t \boldsymbol{S} = 2\mu_B \boldsymbol{S} \times \boldsymbol{B}. \tag{312}$$

The expectation value $\boldsymbol{a} = 2\langle \boldsymbol{S} \rangle$ satisfies the equation of motion

$$\partial_t \boldsymbol{a} = 2\mu_B \boldsymbol{a} \times \boldsymbol{B} \tag{313}$$

and performs a precession around the magnetic field. The equation of motion for the spatial motion in the semi-classical limit,

$$\partial_t \langle \boldsymbol{v} \rangle = \frac{e}{mc} \langle \boldsymbol{v} \rangle \times \boldsymbol{B}, \tag{314}$$

shows that the velocity perform a precession with the same frequency and the angle of $\langle \boldsymbol{S} \rangle$ and $\langle \boldsymbol{v} \rangle$ is time-independent.

To find the generalization for the relativistic case we assume that the right hand side of the equation of motion is linear in $F_{\mu\nu}$, a^μ and at most quadratic in the four velocity, $u^\mu = p^\mu/m$. The invariance of the equation of motion under spatial inversion prevents the appearance of $F^{\mu\nu}u_\nu$ in the equation of motion, the anti-symmetry, $F^{\mu\nu} = -F^{\nu\mu}$, eliminates the term $a^\mu F^{\nu\rho}u_\nu u_\rho$ and the orthogonality $au = 0$ excludes the combination $F^{\mu\nu}u_\nu au$. Therefore one arrives at the equation

$$\partial_s a^\mu = \alpha F^{\mu\nu} a_\nu + \beta u^\mu F^{\nu\rho} u_\nu a_\rho, \tag{315}$$

containing the constants α and β. These parameters can be found by inspecting the equation in the rest frame, in the limit $a^\mu \to (0, \boldsymbol{a})$, $u^\mu \to (1, \boldsymbol{0})$ and $s \to t$,

$$\partial_t \boldsymbol{a} = \alpha \boldsymbol{a} \times \boldsymbol{B}. \tag{316}$$

This result imposes $\alpha = 2\mu_B$. The other parameter is determined by the help of the semi-classical equation of motion,

$$\dot{u}^\mu = \frac{e}{m} F^{\mu\nu} u_\nu, \tag{317}$$

yielding

$$u\dot{a} = -\dot{u}a = -\frac{e}{m} a_\mu F^{\mu\nu} u_\nu \tag{318}$$

where the time derivative of the orthogonality relation, $au = 0$, $\dot{a}u + a\dot{u} = 0$ was used in the first equation. Finally eq. (315) is multiplied by u and we arrive at the equation

$$u\dot{a} = (2\mu_B + \beta) u_\mu F^{\mu\nu} a_\nu = -\frac{e}{m} a_\mu F^{\mu\nu} u_\nu \tag{319}$$

by the help of $u^2 = 1$ which gives

$$\beta = \frac{e}{m} - 2\mu_B \tag{320}$$

and the relativistic equation of motion

$$\partial_s a^\mu = 2\mu F^{\mu\nu} a_\nu + \left(\frac{e}{m} - 2\mu_B\right) u^\mu F^{\nu\rho} u_\nu a_\rho. \tag{321}$$

Appendix A: Multi-valued wave functions

Schrödinger's wave function, a complex valued function of the space-time coordinates, may be multi-valued. It is always an exciting question whether a mathematical possibility is realized by Nature of remains a dead end street of the formalism. In this case the multi-valued nature of the wave function is related to observable, physical phenomena, of which the Aharonov-Bohm effect is the simplest. This is a surprising mechanism of quantum mechanics which generates a new quantum number, without any classical analogy, and an unexpected dependence of expectation values on this quantum number which may lead to the violation of eqs. (27).

1. Particle on the circle

Let us consider a particle which is restricted into a circle of radius r. The requirement that the momentum operator, $p = \frac{\hbar}{ir}\partial_\phi$, be Hermitean brings a restriction for the wave functions. In fact, the matrix elements of the equation $p^\dagger = p$,

$$\langle\psi_1|p|\psi_2\rangle = \int_{-\pi}^{\pi} d\phi \psi_1^*(\phi) \frac{\hbar}{ir} \partial_\phi \psi_2(\phi) = \int_{-\pi}^{\pi} d\phi \left(\frac{\hbar}{ir}\partial_\phi \psi_1(\phi)\right)^* \psi_2(\phi) = \langle\psi_1|p^\dagger|\psi_2\rangle, \tag{A1}$$

assumes the vanishing of the boundary contribution of the partial integration in the second equation which can be assured by imposing the boundary condition,

$$\psi(\phi + 2\pi) = e^{i\theta}\psi(\phi), \tag{A2}$$

on the wave functions of the Hilbert space, $\mathcal{H}_\theta$.

When the momentum operator and the free Hamiltonian, $H = p^2/2m$, are restricted into $\mathcal{H}_\theta$ then its eigenstates are

$$\psi_n(\phi) = e^{i(n+\frac{\theta}{2\pi})\phi} \tag{A3}$$

with the eigenvalues

$$p_\theta\psi = \frac{\hbar}{r}\left(n + \frac{\theta}{2\pi}\right)\psi. \tag{A4}$$

Hence the free particle Hamiltonian has the spectrum

$$H_\theta\psi_n = \frac{\hbar^2}{2mr^2}\left(n + \frac{\theta}{2\pi}\right)^2\psi_n. \tag{A5}$$

The θ-dependence is periodic and the eigenstate are simply shifted, $\psi_n(\phi) \to \psi_{n+1}(\phi)$ as $\theta \to \theta+2\pi$.

It is easy to understand the physical origin of the θ-dependence: We see that the particle return to the same position after a turn around the circle but this fix the wave function up to a phase only and (A2) follows. In other words, the particle interferes with itself as it turns around the circle and the phase difference, $e^{i\theta}$, influences the state and thereby the expectation values, too. There is no classical analogy of θ.

2. Charged particle in a ring, the Aharonov-Bohm effect

We consider now a more realistic problem which turns out to be equivalent with the precedent one. It is about a charge moving on the ring, $\mathcal{R}$, of radius r and negligible thickness which is embedded into the three dimensional space. Furthermore, there a static magnetic field, $\boldsymbol{B} = \boldsymbol{\nabla}\times\boldsymbol{A}$, which is vanishing along the ring and represents a magnetic flux Φ for a surface Σ which is bounded by the ring,

$$\Phi = \int_\Sigma d\boldsymbol{n}\boldsymbol{B}(\boldsymbol{x}) = \oint_{\mathcal{R}} d\boldsymbol{x}\boldsymbol{A}(\boldsymbol{x}). \tag{A6}$$

We use cylindrical coordinates, (z, ρ, ϕ), where the ring corresponds to $z = 0$ and $\rho = r$ and the vector potential, $\boldsymbol{A} = (A_r, A_z, A_\phi)$, is chosen to be $A_z = A_r = 0$ and $A_\phi = \Phi/2\pi r$, a pure gauge

potential, $\boldsymbol{A} = \boldsymbol{\nabla}\alpha(\boldsymbol{x})$ with $\alpha = \phi\Phi/2\pi$, in the vicinity of the ring. The wave functions of the charge satisfy the traditional periodic boundary condition, $\psi(\phi+2\pi) = \psi(\phi)$, and the Hamiltonian, obtained from the free case by the replacement $\boldsymbol{p} \to \boldsymbol{p} - \frac{e}{c}\boldsymbol{A}$ is

$$H_\Phi = \frac{\hbar^2}{2mr^2}\left(\frac{1}{i}\partial_\phi - \frac{e\Phi}{2\pi\hbar c}\right)^2. \tag{A7}$$

The relation with the previous problem should be clear by now: The momentum operator and the Hamiltonian act in $\mathcal{H}_0$ but their spectrum is the one in $\mathcal{H}_{-\frac{e}{\hbar c}\Phi}$, c.f. (A5).

How did we end up with the spectrum of H_θ, working in the Hilbert space $\mathcal{H}_0$? The answer is a gauge transformation,

$$\boldsymbol{A}(\boldsymbol{x}) \to \boldsymbol{A}(\boldsymbol{x}) + \frac{\hbar c}{e}\boldsymbol{\nabla}\chi(\boldsymbol{x}), \quad \psi(\boldsymbol{x}) \to e^{i\chi(\boldsymbol{x})}\psi(\boldsymbol{x}), \tag{A8}$$

with $\chi = -\phi\frac{e}{\hbar c}\frac{\Phi}{2\pi}$ which cancels the vector potential on the ring. This transformation preserves the spectrum of the observables and performs the mapping $\mathcal{H}_\theta \to \mathcal{H}_{\theta-\frac{e}{\hbar c}\Phi}$ and $H_\theta \to H_{\theta-\frac{e}{\hbar c}\Phi}$. In particular, it brings the Hamiltonian (A7), acting on $\mathcal{H}_0$, into the free Hamiltonian, $H_{-\frac{e}{\hbar c}\Phi}$, on the Hilbert space $\mathcal{H}_{-\frac{e}{\hbar c}\Phi}$. The multi-valued gauge transformations, $\frac{e}{\hbar c}\Phi \neq 2\pi n$, map the dynamical system, formulated by the multi-valued wave functions into another system which is described by single-valued wave functions.

The dependence of the expectation values on the magnetic field, called the Aharonov-Bohm effect, is a surprising result on two counts. The first surprise is that the dynamics of the charge is influenced by the magnetic field even this latter is vanishing in the region where the particle propagates. The resolution of this apparent puzzle is that the magnetic field is sufficient to describe the electromagnetic interactions in a gauge invariant scheme only, the gauge non-invariant quantities, such as the wave functions, require the use of the vector potential. The second surprise is that the physical effects are there even if the vector potential in question can be eliminated by an appropriate gauge transformation. The Aharamov-Bohm effect can properly be understood only by copparing two different gauges. On the one hand, it is due to the vector potential in the gauge where the wave functions are periodic. On the other hand, a non-periodic gauge transformation can eliminate the vector potential but the physics remains the same, it is now described by the wave functions with non-trivial boundary conditions.

It is easy to see that the Aharamov-Bohm effect, the interference of the particle with itself is governed by the magnetic flux, a gauge invariant concept. In fact, let us make a full, clockwise turn, $\phi \to \phi - 2\pi$, with the particle along the ring. This rotation is generated by the gauge invariant

momentum operator, $P = \frac{\hbar}{ir}\partial_\phi - eA_\phi$, and it indeed provides the phase factor,

$$\psi(\phi) \to e^{\frac{i}{\hbar}2\pi r P}\psi(\phi) = e^{i\theta}\psi(\phi) = \psi(\phi + 2\pi), \tag{A9}$$

which defines $\mathcal{H}_\theta$.

The representation of discrete symmetries is non-trivial on a Hilbert space with multi-valued wave function and the classical relations like (27) may change. Consider space inversion, $P : \boldsymbol{x} \to -\boldsymbol{x}$, as an example. The transformation of the coordinates in classical physics, $P^2 : \boldsymbol{x} \to \boldsymbol{x}$, suggests $P^2 = \mathbb{1}$ in quantum mechanics. The multi-valued wave functions possess several Riemann-sheets, $\psi_\alpha(\boldsymbol{x})$ with $\psi_\alpha(\boldsymbol{x}) = e^{i\theta_{\alpha,\beta}}\psi_\beta(\boldsymbol{x})$, and the second inversion, performed after the first, may bring us back to a different Riemann-sheet and $P^2 = \mathbb{1}e^{i\theta}$ is possible with $\theta \neq 2n\pi$. In fact, if the first space inversion takes us to the diametrical opposite point along the ring clockwise or anti-clockwise manner and the second will bring us back to the starting position in either of these two ways, too. If the orientation is the same for both inversion then a full turn is realized by P^2 and a non-trivial phase arises, owing to the multi-valuedness of the wave-functions.

3. Dynamics and multi-valued wave functions

A necessary condition for a complex function, $f(z)$, to support several Riemann-sheets and multi-valuedness is an essential singularity somewhere on the complex plane, where the Riemann-sheets collapse to a single value, e.g. $f(z) = \sqrt{z}$ at $z = 0$. To avoid the diverging contribution of the singularity to the kinetic energy the wave function must tend to zero sufficiently fast as the essential singular point is approached in three-space.

Let us consider a free three dimensional particle, in a state defined by the single-valued wave function, $\psi_0(z, \rho, \phi)$, given in the cylindrical coordinate system, $(x, y, x) = (\rho\cos\phi, \rho\sin\phi, z)$. The wave function with multi-valued structure with respect to rotation around the z-axis by 2π, $\psi_\theta(z, \rho, \phi + 2\pi) = e^{i\theta}\psi_\theta(z, \rho, \phi)$, can be written in the form

$$\psi_\theta(z, \rho, \phi) = e^{i\frac{\theta}{2\pi}\phi}\psi_0(z, \rho, \phi). \tag{A10}$$

The expectation value of the kinetic energy,

$$H = -\frac{\hbar^2}{2m}\left(\partial_z^2 + \frac{1}{\rho}\partial_\rho\rho\partial_\rho + \frac{1}{\rho^2}\partial_\phi^2\right), \tag{A11}$$

reveals the characteristic role of the three coordinates: ϕ, is the coordinate of the loop with multi-valued structure, ρ parametrizes the transverse structure of the tube, the particle wih ϕ-dependent

wave function is excluded and finally z is a regular coordinate, without connection with the multi-valued structure. To calculate the kinetice energy we start with the identity $[\partial_\phi, f(\phi)] = f'(\phi)$, to find

$$[\partial_\phi^2, f(\phi)] = \partial_\phi f'(\phi) + f'(\phi)\partial_\phi = f''(\phi) + 2f'(\phi)\partial_\phi \tag{A12}$$

(by the help of $[AB, C] = A[B, C] + [A, C]B$). Hence the θ-dependence of the kinetic energy is an additive term,

$$\langle\psi_\theta|H|\psi_\theta\rangle = \langle\psi_0|H|\psi_0\rangle - \frac{\hbar^2}{2m}\int_{-\infty}^{\infty} dz \int_{-\pi}^{\pi} d\phi \int_0^\infty \frac{d\rho}{\rho}\psi_0^*(z,\rho,\phi)\left[i\frac{\theta}{\pi}\partial_\phi - \left(\frac{\theta}{2\pi}\right)^2\right]\psi_0(z,\rho,\phi). \tag{A13}$$

with $\langle\psi_0|H|\psi_0\rangle \geq 0$ shows that the vanishing of the wave function along the z axes, the exclusion of the particle from $\rho = 0$ is needed to assure the finiteness of the energy. The general rule is that the particle must be excluded from some part of the space to support multi-valued wave functions for rotation by 2π. The remaining, allowed space region where the wave function has finite, non-vanishing magnitude must be multiply-connected, meaning that not all closed loops of the coordinate space can be contracted to a point in a smooth, continuous manner, without passing through excluded regions.

4. Topological symmetry

The quantum mechanics over a topologically non-trivial coordinate space has symmetry property without classical analogy. This is based on the homotopy group of the coordinate space, Q, defined in the following manner. Let us first consider the set of continuous loops, $\gamma : [0, 1] \to Q$, based on a given coordinate, x_0, $\gamma(0) = \gamma(1) = x_0$. One can define an equivalence classes of loops, $\{\Gamma(x_0)\}$, two loop being equivalent, $\gamma_1 \sim \gamma_2$ or $\gamma_1, \gamma_2 \in \Gamma(x_0)$, if γ_1 can continuously be deformed into γ_2. Such an equivalence is expressed by the existence of a continuous function, $f : [0, 1] \otimes [0, 1] \to Q$, with the property $f(s, 0) = \gamma_1(s)$ and $f(s, 1) = \gamma_2(s)$, the parameter t describing the family of the deformed loops, $f(s, t)$. One introduces a group multiplication for the loops by following on loop after the other,

$$\gamma_2 o \gamma_1(s) = \begin{cases} \gamma_1(2s) & 0 < s < \frac{1}{2}, \\ \gamma_2(2s-1) & \frac{1}{2} < s < 1. \end{cases} \tag{A14}$$

It is easy to see that this induces a multiplication for the equivalence classes, $\Gamma_2 o \Gamma_1$ consisting of the loops $\gamma_2 o \gamma_1$, $\gamma_j \in \Gamma_j$. The group, obtained in such a manner, is called the first homotopy group,

$\pi_1(Q)$. It is independent of the choice of the base point, x_0. Note that the transport of the system over a closed loop is a true symmetry, having no ways to distinguish the system before and after the transport and $\pi_1(Q)$ is a symmetry group.

The find the physical role of the homotopy group one regards the transport of a system along a closed loop as a symmetry transformation since the state of the system before and after the transport are identical. The wave functions are defined up to a global phase hence the transport may induce such a path-dependent phase, $e^{i\theta}$. Let us suppose that the wave function supports a multi-valued structure in a compact coordinate and start with a Hilbert space, $\mathcal{H}_0$, consisting of single-valued wave functions. The multi-valued, extended Hilbert space, $\mathcal{H}_\theta$, can be constructed in a manner, similar to the aperiodic gauge transformations, mentioned above. The highly non-trivial point is the relevance of topology, i.e. that the phase factor, $e^{i\theta}$, is preserved by the continuous deformation of the loop. (This can easily be seen in the path integral representation of the transition amplitudes in the following manner: (i) The Schrödinger equation can be derived from the path integral expressions by performing infinitesimal variation of the paths. This shows that the minial functional space to extend the path integration must be closed with respect to fininitesimal variations. The path integral is well defined within this functional space. (ii) We may assign arbitraty phase factors to the path integral within disconnected subspaces of the trajectory space, if they exist, without modifying the Schrödinger equation. (iii) The trajectories, belonging to different homotopy classes of the coordinate space represent trajectories, within disconnected path families.) The phases, arising from the multi-valuedness, are chosen in such a manner that they induce a representation of the homotopy group. The impact of this symmetry on the dynamics is that the particle states before and after the transport interferes and their relative phase being $\theta(\Gamma)$. Julian Schwinger expressed this is by saying that while the state of a classical particle is characterized by its location in the phase space the characterization of a quantum particle requires not only the knowledge the actual state but also the way how it ended up there. The Aharonov-Bohm effect is based on the additive group of integers, $\pi_1(U(1)) = Z$, called winding number,

$$\nu(\gamma) = \int_0^{2\pi} \frac{d\phi}{2\pi} \dot{\gamma}(\phi), \tag{A15}$$

and the boson-fermion classification originates from $\pi_1(SO(3)) = Z_2 = \{\mathbb{1}, -\mathbb{1}\}$.