Motion in External Electromagnetic Field. Gauge Fields in Quantum Mechanics.

Selected chapter of lecture notes on Quantum Mechanics

S. Levit
Dept. of Condensed Matter
The Weizmann Institute of Science
Rehovot, Israel

December 3, 2024

Contents

Electromagnetic potentials $\mathbf{A}(\mathbf{r},t)$ and $A_0(\mathbf{r},t)$ appear in classical physics as auxiliary quantities which are introduced in order to simplify the form and solutions of the Maxwell equations, cf., Chapter 10 in Ref. [1]. The fact that they are not uniquely defined and can be changed without affecting any physical results by a transformation bearing a strange name of "gauge" seems to be rather an annoying nuisance than a fundamental symmetry of nature.

This state of affairs undergoes drastic revision when quantum mechanical description is attempted. We do not know how to formulate such a description in the presence of the electromagnetic field without making an essential use of the electromagnetic potentials. Moreover the invariance under the gauge transformations becomes a profound symmetry of our world which lies at the origin of all known interactions. Because of this the fields which carry these interactions are termed *gauge fields*.

The problem of the quantum mechanical motion in an external electromagnetic field provides the simplest setup in which one encounters some of the strange and beautiful phenomena appearing as a result of the symbiosis of gauge fields and quantum mechanics.

Note. I have changed from CGS to SI units in Sections 1-8. The rest is in CGS.

1 Electromagnetic Potentials. The Hamiltonian

1.1 Electromagnetic potentials in classical physics

Let us begin by briefly recalling how the electromagnetic potentials are introduced. Classical electromagnetic field is described by two vector fields $\mathbf{E}(\mathbf{r},\mathrm{t})$ and $\mathbf{B}(\mathbf{r},\mathrm{t})$. In the present chapter these fields will be considered as external, i.e. produced by sources (electric charges and currents) which dynamically are not a part of the physical system under consideration and are not effected by it. This means that the back reaction of the system on the sources of the field is negligible. Although in such circumstances $\mathbf{E}$ and $\mathbf{B}$ should be regarded as controlled externally by charge and current distributions $\rho(\mathbf{r},t)$ and $\mathbf{j}(\mathbf{r},t)$ of the sources via

$$\nabla\cdot\mathbf{E}=\frac{\rho}{\epsilon_0}\quad,\quad \nabla\times\mathbf{B}=\mu_0\mathbf{j}+\frac{1}{c^2}\frac{\partial\mathbf{E}}{\partial t}$$

they can not be taken as completely arbitrary. Indeed irrespective of the configuration of ρ and $\mathbf{j}$ these fields must satisfy the homogeneous pair of Maxwell equations

$$\nabla \cdot \mathbf{B} = 0 \quad , \quad \nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} \tag{1}$$

at every point in space and time. In order to have these equations automatically satisfied the familiar vector and scalar potentials $\mathbf{A}(\mathbf{r},\mathrm{t})$ and $A_0(\mathbf{r},\mathrm{t})$ are introduced[1]. This is done by noticing that the first of the equations above means that $\mathbf{B}$ must be a curl of a vector field $\mathbf{A}(\mathbf{r},t)$. Using this in the second equation gives

$$\nabla \times (\mathbf{E} + \frac{\partial \mathbf{A}}{\partial t}) = 0$$

restricting the combination $\mathbf{E} + \partial\mathbf{A}/\partial t$ to be a gradient of a scalar field. One has therefore

$$\begin{aligned} \mathbf{E} &= -\frac{\partial \mathbf{A}}{\partial t} - \nabla A_0 \ , \\ \mathbf{B} &= \nabla \times \mathbf{A} \end{aligned} \tag{2}$$

Unlike the field strengths $\mathbf{E}$ and $\mathbf{B}$, the electromagnetic potentials can be regarded as unrestricted so that any $\mathbf{A}(\mathbf{r},t)$ and $A_0(\mathbf{r},t)$ can be realised by the poper choice of the external charge and current distributions.

The use of the electromagnetic potentials however presents another problem. They are not unique since the gauge transformation

$$\begin{aligned} \mathbf{A}'(\mathbf{r},\mathrm{t}) &= \mathbf{A}(\mathbf{r},\mathrm{t}) + \nabla\chi(\mathbf{r},\mathrm{t}) \\ A_0'(\mathbf{r},\mathrm{t}) &= A_0(\mathbf{r},\mathrm{t}) - \frac{\partial\chi(\mathbf{r},\mathrm{t})}{\partial t} \end{aligned} \tag{3}$$

with an arbitrary function $\chi(\mathbf{r},\mathrm{t})$ leaves $\mathbf{E}$ and $\mathbf{B}$ invariant. As was already mentioned above this invariance, called the gauge invariance, has profound consequences in quantum mechanical systems and will be discussed at length below. At the moment we just notice that because of it only three among the four functions $\mathbf{A}$ and A_0 are independent. In general one combination of the four functions can be eliminated by a suitably chosen gauge transformation. For instance choosing

$$\chi(\mathbf{r},\mathrm{t}) = \int_{t_0}^{t} A_0(\mathbf{r},t')\,dt'$$

(with arbitrary t_0) eliminates A_0 and leaves $\mathbf{A}(\mathbf{r},\mathrm{t})$ as the only independent degrees of freedom of the electromagnetic field.

[1] Although we use "relativistic" notation for A_0 we use "non relativistic" terminology and call it a scalar potential

1.2 Classical Hamiltonian and equations of motion

Classical non relativistic equation of motion for a particle with electric charge q and mass m in a given electromagnetic field is obtained by using the Lorenz force in the Newton law

$$m\frac{d^2\mathbf{r}}{dt^2} = q\mathbf{E} + q\left(\frac{d\mathbf{r}}{dt}\times\mathbf{B}\right) \ . \tag{4}$$

In order to obtain the quantum mechanical description one can follow either the canonical or the path integral quantization procedures. We will start with the former. We first determine the classical canonical variables and the classical Hamiltonian function of the problem.

The above equation is in terms of coordinates $\mathbf{r}(t)$ and velocities $\mathbf{v}(t) = d\mathbf{r}/dt$ so it is is most convenient to start by determining the Lagrangian of the system. This is

$$L(\mathbf{r},\mathbf{v},t) = \frac{1}{2}m\mathbf{v}^2 + q\mathbf{A}\cdot\mathbf{v} - qA_0(\mathbf{r}) \ . \tag{5}$$

Indeed have

$$\frac{d}{dt}\frac{\partial L}{\partial \mathbf{v}} = m\frac{d\mathbf{v}}{dt} + q\frac{d\mathbf{A}}{dt} = m\frac{d\mathbf{v}}{dt} + q\frac{\partial \mathbf{A}}{\partial r_j}\frac{dr_j}{dt} + q\frac{\partial \mathbf{A}}{\partial t}$$

and

$$\frac{\partial L}{\partial \mathbf{r}} = q\frac{\partial}{\partial \mathbf{r}}(\mathbf{A}\cdot\mathbf{v}) - q\frac{\partial A_0}{\partial \mathbf{r}}$$

In components

$$m\frac{dv_i}{dt} + q\frac{\partial A_i}{\partial r_j}v_j + q\frac{\partial A_i}{\partial t} = q\frac{\partial A_j}{\partial r_i}v_j - q\frac{\partial A_0}{\partial r_i}$$

So have

$$m\frac{dv_i}{dt} = q\left(-\frac{\partial A_i}{\partial t} - \frac{\partial A_0}{\partial r_i}\right) + q\left(\frac{\partial A_j}{\partial r_i} - \frac{\partial A_i}{\partial r_j}\right)v_j$$

which is the Newton equation (4). Indeed recalling Eq.(2) one sees that the fist term is $q\mathbf{E}$, while the last term can be transformed as

$$\epsilon_{ijk}v_jB_k = (\mathbf{v}\times\mathbf{B})_i$$

where we used the antisymmetric symbol ϵ_{ijk}[1] to write vector products, e.g

$$(\mathbf{C}\times\mathbf{D})_i = \epsilon_{ijk}C_jD_k \ \ , \ \ C_iD_j - C_jD_i = \epsilon_{ijk}(\mathbf{C}\times\mathbf{D})_k \ .$$

[1]The Levi–Civita symbol ϵ_{ijk} is defined by $\epsilon_{123} = 1$ and the antisymmetry property under interchange of any indices, $\epsilon_{ijk} = -\epsilon_{jik} = -\epsilon_{ikj}$, etc. ϵ_{ijk} does not change under cyclic permutations $\epsilon_{ijk} = \epsilon_{kij} = ...$.

These two equalities are related by a useful identity

$$\epsilon_{ijk}\epsilon_{ij'k'} = \delta_{jj'}\delta_{kk'} - \delta_{jk'}\delta_{j'k} \ .$$

The above calculations show that the canonical momentum is

$$\mathbf{p} = \frac{\partial L}{\partial \mathbf{v}} = m\frac{d\mathbf{r}}{dt} + q\mathbf{A}(\mathbf{r}) \ . \tag{6}$$

which expresses perhaps the most unusual aspect of the motion in the EM field - the fact that $\mathbf{p} \neq m\mathbf{v}$. In the literature one often meets the term "kinetic momentum" referring to the familiar $m\mathbf{v}$.

Expressing $\mathbf{v}(\mathbf{p},\mathbf{r}) = (\mathbf{p} - q\mathbf{A}(\mathbf{r}))/m$ and using

$$H = \mathbf{p}\cdot\mathbf{v} - L$$

with the above $\mathbf{v}(\mathbf{p},\mathbf{r})$ we find the Hamiltonian function

$$H(\mathbf{p},\mathbf{r}) = \frac{1}{2m}\left(\mathbf{p} - q\mathbf{A}(\mathbf{r},\mathrm{t})\right)^2 + qA_0(\mathbf{r},\mathrm{t}) \ , \tag{7}$$

It is not difficult (and not surprising) to show that with this $H(\mathbf{p},\mathbf{r})$ the equation of motion (4) is equivalent to the two Hamilton equations

$$\frac{d\mathbf{r}}{dt} = \frac{\partial H}{\partial \mathbf{p}} \quad , \quad \frac{d\mathbf{p}}{dt} = -\frac{\partial H}{\partial \mathbf{r}}$$

2 Quantization

2.1 The orbital part

Having established the form of H we follow the canonical quantization procedure and consider the Schrödinger equation with the Hamiltonian operator which is obtained by replacing $\mathbf{r}$ and $\mathbf{p}$ in H by the operators $\mathbf{r}_{op} = \mathbf{r}$ and $\mathbf{p}_{op} = -i\hbar\nabla$,

$$H_{op} = \frac{1}{2m}\left[-i\hbar\nabla - q\mathbf{A}(\mathbf{r},t)\right]^2 + qA_0(\mathbf{r},t) \ . \tag{8}$$

2.2 The spin magnetic moment

Experimental evidence shows that this Hamiltonian is capable of describing only particles which do not carry spin. It must be modified when the spin degrees of freedom are present. This should not be too surprising since already in classical physics the energy of a spinning *charged* particle receives an additional contribution apart from the orbital motion. This contribution arises from the interaction with the magnetic field $\mathbf{B}$ of a localized distribution of electric current $\mathbf{j}(\mathbf{r})$ which a spinning charge creates. For a "point like" particle, i.e. a particle the size of which is much smaller than the scale over which $\mathbf{B}(\mathbf{r})$ changes, the corresponding energy is

$$E_{\text{spinning charge}} = -\boldsymbol{\mu} \cdot \mathbf{B}(\mathbf{r})$$

where $\boldsymbol{\mu}$ is the magnetic moment of the current, cf., Chapter 5 of the Ref. [1],

$$\boldsymbol{\mu} = \frac{1}{2} \int d^3r \; \mathbf{r} \times \mathbf{j}(\mathbf{r}) \; .$$

For composite particles the total current is a sum over internal components

$$\mathbf{j}(\mathbf{r}) = \sum_a q_a \mathbf{v}_a \delta(\mathbf{r} - \mathbf{r}_a)$$

each with its dynamics so the calculation of $\boldsymbol{\mu}$ is in general not an easy task. However if all the charged components have an equal charge to mass ratio $q_1/m_1 = q_2/m_2 = ... = q/m$ the magnetic moment can be written as

$$\boldsymbol{\mu} = \frac{1}{2} \sum_a q_a (\mathbf{r}_a \times \mathbf{v}_a) = \frac{q}{2m} \sum_a \mathbf{r}_a \times m_a \mathbf{v}_a = \frac{q}{2m} \mathbf{L} \; . \tag{9}$$

Experimental data as well as theoretical considerations (cf., Section 8 below) indicate that for elementary particles like electrons this classical linear relation between $\boldsymbol{\mu}$ and the angular momentum of a system holds also between the corresponding quantum mechanical operators of the magnetic moment $\boldsymbol{\mu}_{op}$ and the spin $\mathbf{s}_{op}$. However the proportionality coefficient in general does not coincide with the classical value. To emphasize this difference it is conventional (for charged particles) to write the relation between the operators $\boldsymbol{\mu}$ and $\mathbf{s}$ as

$$\boldsymbol{\mu}_{op} = g \; \frac{q}{2m} \; \mathbf{s}_{op} \tag{10}$$

with q - the particle charge and g - dimensionless coefficient called the gyromagnetic factor or for short the g-factor. Theoretical methods which allow to determine g and examples of their applications are considered in Section 8.

2.3 The Schrödinger equaion

Adding the term $-\boldsymbol{\mu}_{op} \cdot \mathbf{B}$ to the Hamiltonian operator (8) one can write the Hamiltonian for an elementary particle with a spin in an external EM field as

$$H_{op} = \frac{1}{2m}\left(-i\hbar\nabla - q\mathbf{A}\right)^2 + qA_0 - g\frac{q}{2m}\ \mathbf{s}_{op} \cdot \mathbf{B}\ . \tag{11}$$

and the corresponding Schrödinger equation

$$i\hbar\frac{\partial\psi}{\partial t} = \left[\frac{1}{2m}\left(-i\hbar\nabla - q\mathbf{A}\right)^2 + qA_0 - g\frac{q}{2m}\ \mathbf{s}_{op} \cdot \mathbf{B}\right]\psi \tag{12}$$

where

$$\psi = \psi(\mathbf{r}, \sigma; t)$$

is a function of space and spin variables $\mathbf{r}$ and σ.

In writing out the square in this equation one should not forget that the operator $\mathbf{p}_{op} = -i\hbar\nabla$ in general does not commute with the vector $\mathbf{A}$ which is a function of coordinates. Since $\mathbf{p}_{op} \cdot \mathbf{A} - \mathbf{A} \cdot \mathbf{p}_{op} = -i\hbar\nabla \cdot \mathbf{A}$, one can write

$$\frac{1}{2m}\left(-i\hbar\nabla - q\mathbf{A}\right)^2 = -\frac{\hbar^2}{2m}\nabla^2 + \frac{i\hbar q}{2m}(\nabla \cdot \mathbf{A} + 2\mathbf{A} \cdot \nabla) + \frac{q^2}{2m}\mathbf{A}^2\ .$$

The operators $\mathbf{p}_{op}$ and $\mathbf{A}$ commute if $\nabla \cdot \mathbf{A} = 0$. This happens e.g., for $\mathbf{A} = (\mathbf{B} \times \mathbf{r})/2$ which is a possible choice of $\mathbf{A}$ in a particular case of a uniform magnetic field[2].

In the following sections we will examine various properties of the equation (12) and will present its solutions for some particular simple choices of the electric and magnetic fields.

3 Gauge Invariance

3.1 Gauge transformations in quantum mechanics

The electromagnetic field enters the classical and quantum equations (4) and (12) via very different sets of variables. The classical equation depends on the physically

[2] Verifying

$$\begin{aligned}(\nabla \times \mathbf{A})_i &= \frac{1}{2}\epsilon_{ijk}\nabla_j\epsilon_{klm}B_l x_m = \frac{1}{2}(\delta_{il}\delta_{jm} - \delta_{im}\delta_{jl})\delta_{jm}B_l = \frac{1}{2}(3B_i - B_i) = B_i \\ \nabla \cdot \mathbf{A} &= \nabla_i A_i = \nabla_i\epsilon_{ijk}B_j x_k = \epsilon_{ijk}B_j\delta_{ik} = 0\end{aligned} \tag{13}$$

measurable variables **E** and **B** of the field whereas in the Schrödinger equation the field enters via non uniquely defined and seemingly auxiliary objects **A** and A_0. This is not an accident. At present no formulation of quantum mechanics exists which does not explicitly use the electromagnetic potentials. Schrödinger and Heisenberg pictures require the Hamiltonian while the path integral quantization uses the Lagrangian (cf., below, Section 10) and both objects can not be written without **A** and A_0. Since the potentials are not uniquely defined and can be changed by a gauge transformation one must address the question of how unambiguous physical results are obtained in such a situation.

Unlike in classical mechanics where gauge transformations do not change the equations of motion the Schrödinger equation (12) and therefore also its solutions $\psi(\mathbf{r},\mathrm{t})$ are transformed in a non trivial way[3]. It is not difficult to find how the transformation of $\psi(\mathbf{r},\mathrm{t})$ is related to the transformation of the potentials. For this we notice that **A** and A_0 enter the equation only in the combinations

$$-i\hbar\mathbf{D} \equiv (-i\hbar\nabla - q\mathbf{A}) \quad \text{and} \quad i\hbar D_0 \equiv (i\hbar\frac{\partial}{\partial t} - qA_0) \tag{14}$$

so that the Schrödinger equation (12) can be written as

$$i\hbar D_0\psi = -\frac{\hbar^2}{2m}\mathbf{D}^2\psi + \text{spin part} \tag{15}$$

If $\psi(\mathbf{r},\mathrm{t})$ is a solution for a particular choice of **A** and A_0 then

$$\psi(\mathbf{r},\mathrm{t}) = \exp\left[-i\frac{q}{\hbar}\chi(\mathbf{r},\mathrm{t})\right]\psi\,'(\mathbf{r},\mathrm{t}) \equiv S(\mathbf{r},\mathrm{t})\psi\,'(\mathbf{r},\mathrm{t}) \tag{16}$$

satisfies

$$\begin{aligned} -i\hbar\mathbf{D}\psi &= (-i\hbar\nabla - q\mathbf{A})\,\psi = S(\mathbf{r},\mathrm{t})\,(-i\hbar\nabla - q\mathbf{A}')\,\psi' = -i\hbar S(\mathbf{r},\mathrm{t})\mathbf{D}'\psi'\,, \\ i\hbar D_0\psi &= \left(i\hbar\frac{\partial}{\partial t} - qA_0\right)\psi = S(\mathbf{r},\mathrm{t})\left(i\hbar\frac{\partial}{\partial t} - qA_0'\right)\psi\,' = i\hbar S(\mathbf{r},\mathrm{t})D_0'\psi' \end{aligned} \tag{17}$$

and therefore solves the Schrödinger equation for the transformed potentials, Eq. (3) (please note the primed $\mathbf{D}'$ and D_0' on the right hand side of the expressions above).

We see that the classical concept of the gauge transformation undergoes a generalization in quantum mechanics. Now not only the potentials which describe the

[3]In this and many of the following sections the dependence of $\psi(\mathbf{r},\sigma;t)$ on the spin variable σ will not be of interest and will be suppressed for brevity.

electromagnetic field but also the wave functions describing the material particles must change simultaneously according to the rules (3) and (16). This change is *local*, i.e. it is different for different points in space and time. One often emphasizes this aspect by calling the transformation given by Eqs. (3), (16) a *local gauge transformation* to distinguish it from a *global transformation* in which the wave function is multiplied by a constant phase factor.

It is seen that the combinations $\mathbf{D}\psi$ and $D_0\psi$ defined in (17) transform under a local gauge transformation in a particularly simple way – i.e. as if it were a global transformation. These combinations are called gauge covariant derivatives in theories with gauge fields. The way to introduce the electromagnetic field in the dynamical equations by replacing the ordinary derivatives $\partial/\partial t$ and $\partial/\partial\mathbf{r}$ by the gauge covariant combinations $D_0, \mathbf{D}$ is known as *minimal coupling*.

3.2 Gauge symmetry vs gauge invariance

We may now ask a question as to whether the classical gauge invariance also holds in quantum mechanics, namely whether the result of any measurement is invariant under gauge transformations which now include also the local transformation (16) of the wave function. It is an empirical fact that the answer to this question is positive. Moreover it is also clear that this invariance known as *the local gauge invariance* is a profound fundamental symmetry of the quantum mechanical description in the presence of gauge fields.

It is important to note that this symmetry does not mean that the wave functions must be invariant. Like with other fundamental symmetries, e.g. the invariance with respect to translations and rotations, the gauge symmetry means that the wave functions transform in a particular way given by Eq. (16), i.e they form a representation of the corresponding group of transformations.

Here, however, the similarity ends. Unlike other symmetries the gauge symmetry demands that the *observable quantities* must not be effected by the gauge transformations and therefore must be "gauge scalars", i.e. depend on gauge invariant combinations of $\psi, \mathbf{A}$ and A_0. No "gauge vectors", "gauge tensors", etc, are ever observed. The origin of this difference can only be understood when the full quantum dynamics of the electromagnetic field and its coupling to matter are discussed.

We conclude this section by noting that explicit appearance of the electromagnetic potentials in the equations of quantum mechanics makes the gauge invariance a very subtle symmetry. Its consequences and generalizations are important aspects of the modern physics. We will make a special point in this chapter to illustrate some of the related physical ideas and results.

3.3 The Gauge Principle – symmetry dictates interactions

In the previous section we started with the known transformation properties of the potentials and then on the basis of the special manner in which they entered the Schrödinger equations – i.e. in combinations $\mathbf{D}$ and D_0, – derived the required transformation properties of the wave functions which were necessary in order to keep the Schrödinger equation form-invariant.

Imagine now that we reverse this derivation in the following manner. Let us begin by considering the free Schrödinger equation

$$i\hbar\partial_t\psi = -\hbar^2\nabla^2\psi/2m \quad .$$

This equation is obviously invariant under the *global gauge transformations* i.e. the transformations (16) with a constant χ independent of ($\mathbf{r}$,t). This global gauge invariance is a fundamental feature of the Schrödinger equation. One of its notable consequences is the conservation of the integral $\int d\mathbf{r}\ \psi^*(\mathbf{r},\mathrm{t})\psi(\mathbf{r},\mathrm{t})$. This integral is the total probability or, when multiplied by e, the total electric charge. The relation of its conservation to the global gauge invariance is not intuitively obvious but can be rigorously derived by the applications of arguments of the Noether theorem to the Schrödinger field.

Now let us see what happens if one demands that the nature should be invariant not only under the global but also under *local gauge transformations*, i.e. with the ($\mathbf{r}$,t)-dependent phase χ in Eq. (16). It is obvious that the free Schrödinger equation will not satisfy this demand since its derivatives will act on the local phase producing additional terms with $\nabla\chi$ and $\partial\chi/\partial t$. With the hindsight of the previous section we can however write a more general Schrödinger equation which will be locally gauge invariant.

In order to compensate for the derivatives $\nabla\chi$ and $\partial\chi/\partial t$ and eliminate them from the transformed Schrödinger equation we must

(a) "postulate" the existence of a field described by the potentials $\mathbf{A}$ and A_0,

(b) replace the ordinary derivatives $\partial/\partial t$ and ∇ in the equation by the gauge covariant combinations D_0, $\mathbf{D}$ and

(c) require that the potentials transform according to Eq. (3) simultaneously with the transformation (16) of the wave functions.

The demand of the *local gauge invariance* is thus turned into a powerful *heuristic* principle – *The Gauge Principle*, which, had we not known about the electromagnetic field, led us to "discover" its existence and the way it must appear in the Schrödinger equation.

Of course the last, spin-dependent term in (12) would not be deduced in such a procedure and should be justified separately. The need for this separate discussion of the spin interaction with the electromagnetic field disappears when a fully relativistic theory of *elementary particles* is considered, cf. Section 8.1 in Ref. [2] or Chapter 3 in Ref. [3]. Moreover it can be shown that the entire Maxwell electrodynamics is fully consistent with the The Gauge Principle supplemented by very general requirements of the time-space translational invariance and the Lorenz invariance.

It also turns out that the fields responsible for all other known interactions, i.e. weak, strong and gravitational are consistent with The Gauge Principle in a similar way. Namely for every known interaction there exist a a global symmetry of a non interacting theory which becomes a local symmetry after the interaction is introduced. The potentials describing the interaction are the compensating gauge potentials which are necessary to introduce in order to satisfy this demand are the fields of the fundamental interactions. Thus The Gauge Principle essentially means that *Symmetry Dictates Interactions*. The Gauge Principle for general relativity for example means that the theory is invariant under *local* Lorenz transformations. In Section 12 below we consider an example of how a so called non abelian gauge field appears as a result of the demand that the Schrödinger equation is invariant under local non abelian transformations.

4 Electric Current Density.

4.1 The orbital part

Let us derive the quantum mechanical expression for the current density of charged particles. We will start by considering the continuity equation for the charge density $\rho(\mathbf{r},\mathrm{t}) = q\psi^*(\mathbf{r},\mathrm{t})\psi(\mathbf{r},\mathrm{t})$. Multiplying the Schrödinger equation (12) on the left by ψ^* and its complex conjugate by ψ and subtracting one obtains in a standard way that $\partial\rho/\partial t + \nabla\cdot\mathbf{j} = 0$ with the current density

$$\mathbf{j}_{\mathrm{orbital}}(\mathbf{r}) = \frac{q}{2m}\left[\psi(\mathbf{r})(i\hbar\nabla - q\mathbf{A}(\mathbf{r}))\psi^*(\mathbf{r}) + \psi^*(\mathbf{r})(-i\hbar\nabla - q\mathbf{A}(\mathbf{r}))\psi(\mathbf{r})\right] . \qquad (18)$$

Let us note that strictly speaking this expression is correct for a spinless charged particle. To include spin would require to account for the spin dependence of the wavefunction, i.e. to replace $\psi(\mathbf{r}) \to \psi(\mathbf{r},\sigma)$ and to sum over the spin index σ. In the particular case of the space-spin separation

$$\psi(\mathbf{r},\sigma) \to \psi(\mathbf{r})\xi(\sigma)$$

one recovers Eq. (18).

One should also observe that the expression $\mathbf{j}_{\rm orbital}(\mathbf{r})$ for the current is actually the expectation value $<\psi|\mathbf{j}_{op}(\mathbf{r})|\psi>$ of the operator

$$\mathbf{j}_{op}(\mathbf{r}) = \frac{1}{2}\left[q\mathbf{v}_{op}\delta(\mathbf{r}-\mathbf{r}_{op}) + \delta(\mathbf{r}-\mathbf{r}_{op})\, q\mathbf{v}_{op}\right] \tag{19}$$

of the current density due to orbital motion with the velocity

$$\mathbf{v}_{op} = [\mathbf{p}_{op} - q\mathbf{A}(\mathbf{r}_{op})]/m \ .$$

This operator is just what is obtained from the classical expression $\rho(\mathbf{r},t)\mathbf{v}(t) = q\delta(\mathbf{r}-\mathbf{r}(t))\mathbf{v}(t)$ for a point particle by replacing the classical quantities $\mathbf{r}(t)$ and $\mathbf{v}(t)$ with the corresponding operators $\mathbf{r}_{op}$ and $\mathbf{p}_{op}$ and symmetrizing the final expression in order to make it hermitian.

4.2 The spin contribution

The missing feature in the above expression for the current is the absence of the contribution from the spin of the particle. This is the reason we have added to it the index *orbital*. As we have already discussed a spinning charged particle creates a local distribution of electric current at its location and one should expect to find an appropriate term in the current density in addition to the contribution of the orbital motion. We have missed this term because as we will see in a moment it is in the form of a rotor of a vector (a so called solenoidal term) and therefore can not be seen in the continuity equation which depends only upon the divergence of the current.

In order to correct our result let us consider a physical system of charges $\{q_a\}$ placed in positions $\{\mathbf{r}_a\}$ and put it under the influence of an external electric field $\mathbf{E}(\mathbf{r})$.[4] We start classically and consider a time interval dt during which these charges move distances $d\mathbf{r}_a = \mathbf{v}_a dt$. As a result their total energy is changed by

$$dW = \sum_a q_a \mathbf{E}(\mathbf{r}_a)\cdot d\mathbf{r}_a = dt\int d\mathbf{r}[\sum_a q_a \frac{d\mathbf{r}_a}{dt}\delta(\mathbf{r}-\mathbf{r}_a)]\cdot\mathbf{E}(\mathbf{r})$$

The expression in the square brackets here is the total current density flowing in the system, so that

$$\frac{dW}{dt} = \int d\mathbf{r}\ \mathbf{j}(\mathbf{r})\cdot\mathbf{E}(\mathbf{r}) \ . \tag{20}$$

[4] We follow Sec. 114 in Ref. [6]

We assume that this relation holds also for quantum mechanical expectation values. Considering for simplicity one particle and let us form the expectation value of the Hamiltonian (11)

$$\begin{aligned} W &= \langle\psi|H|\psi\rangle \\ &= \langle\psi|\left[\frac{1}{2m}\left(-i\hbar\nabla - q\mathbf{A}\right)^2 + qA_0 - g\frac{q}{2m}\ \mathbf{s}_{op}\cdot\mathbf{B}\right]|\psi\rangle \end{aligned} \tag{21}$$

We also have

$$\frac{dW}{dt} =< \psi|\frac{\partial H}{\partial t}|\psi > \quad . \tag{22}$$

This relation (sometimes called the Feynman–Hellmann theorem) is valid since the term $< \partial\psi/\partial t|H|\psi > + < \psi|H|\partial\psi/\partial t >$ vanishes on account of the Schrödinger equation $i\hbar\,\partial\psi/\partial t = H\psi$.

In order to find the time derivative of the Hamiltonian we note that it depends on time only via the time dependence of the potentials $\mathbf{A}, A_0$. Part of this time dependence is not physical and is related to the time dependent gauge transformations of $\mathbf{A}$ and A_0. In order to avoid this fake time dependence we fix the gauge by choosing $A_0 = 0$. This choice does not fix the potentials completely but the only freedom left is *time independent* gauge transformations, i.e. Eq. (3) with time independent $\chi(\mathbf{r})$. With this choice we have that

$$\frac{\partial H}{\partial t} = \int d\mathbf{r}\ \frac{\delta H}{\delta\mathbf{A}(\mathbf{r},\mathrm{t})}\ \frac{\partial\mathbf{A}(\mathbf{r},\mathrm{t})}{\partial t}\ . \tag{23}$$

Using (2) with $A_0 = 0$ and Eq. (20) we obtain the general relation for the electric current

$$\mathbf{j}(\mathbf{r},\mathrm{t}) = - < \psi|\frac{\delta H}{\delta\mathbf{A}(\mathbf{r},\mathrm{t})}|\psi > \ . \tag{24}$$

Varying H with respect to $\mathbf{A}$ and using $\mathbf{B} = \nabla\times\mathbf{A}$ we obtain

$$\begin{aligned} < \psi|\delta H|\psi > \ &= \ \sum_{spin}\int d\mathbf{r}\left\{\psi^*\left[\frac{iq\hbar}{2m}(\nabla\cdot\delta\mathbf{A} + \delta\mathbf{A}\cdot\nabla) + \frac{q^2}{m}\delta\mathbf{A}\cdot\mathbf{A}\right]\psi\right. \\ &\left. - g\frac{q}{2m}(\psi^*\mathbf{s}_{op}\psi)\cdot(\nabla\times\delta\mathbf{A})\right\} \end{aligned} \tag{25}$$

where $\sum_{spin}$ stand for the summation over the spin indices in $\psi(\mathbf{r},\sigma)$.

Integrating by parts in the first term, using the identity

$$\mathbf{a}\cdot\nabla\times\mathbf{b} = -\nabla\cdot(\mathbf{a}\times\mathbf{b}) + \mathbf{b}\cdot(\nabla\times\mathbf{a})$$

for the last term in this expression and assuming that the surface terms vanish we obtain the following expression for the current

$$\mathbf{j}(\mathbf{r}) = \frac{iq\hbar}{2m}(\psi\nabla\psi^* - \psi^*\nabla\psi) - \frac{q^2}{m}\mathbf{A}\psi\psi^* + g\frac{q}{2m}\nabla\times(\sum_{spin}\psi^*\mathbf{s}_{op}\psi) \tag{26}$$

The first two terms are just the "orbital" current already obtained earlier from the continuity equation. For brevity we have omitted the $\sum_{spin}$ in this part as trivially diagonal in spin indices.

The last, "solenoidal" term is the spin contribution which has the appearance of the classical relation between the current and magnetization, i.e. the magnetic moment density $\mathbf{m}(\mathbf{r})$

$$\mathbf{j}_{spin}(\mathbf{r}) = \nabla\times\mathbf{m}(\mathbf{r}) \quad , \quad \mathbf{m}(\mathbf{r}) \equiv \sum_{\sigma'\sigma}\psi^*(\mathbf{r},\sigma')\mathbf{s}_{\sigma'\sigma}\psi(\mathbf{r},\sigma)$$

4.3 Convective, diamagnetic and spin parts of the current

The first term in the expression (26) for the current $\mathbf{j}$ is called the convection current and coincides with the usual expression for the current density in the absence of the electromagnetic field. It is not gauge invariant without the second term which is called the diamagnetic current. The third, spin term in $\mathbf{j}$ is obviously gauge invariant by itself.

In elementary quantum mechanics one develops certain intuition about currents associated with given wave functions. In particular one is used to the fact that non vanishing current density does not appear if the wave function is real, that the current is related to the local complex phase of ψ, etc. This intuition is founded entirely on the first term in Eq. (26) and could be misleading in the presence of electromagnetic field. In this case one finds for instance a non vanishing orbital current density

$$\mathbf{j}_{\rm orbital}(\mathbf{r},{\rm t}) = (q^2/m)\mathbf{A}\psi^2(\mathbf{r}) \ \ ({\rm for} \ \ \psi - {\rm real})$$

for a real wave function. Of course the freedom of local gauge transformations (16) makes the phase of ψ and the difference between real and complex wave functions into something which depends on the choice of the gauge and therefore unphysical.

5 Motion in a Uniform Electric Field

Already such a simple problem as the motion of a charged particle in a constant uniform electric field $\mathbf{E}$ exhibits peculiarities of gauge fields in quantum mechanics.

Classically everything is simple. The particle moves with the constant acceleration $q \mid \mathbf{E} \mid /m$ in the direction of the field and has a constant, determined by initial conditions velocity perpendicular to this direction. In quantum mechanics one may have differently looking descriptions depending on which of the many (i.e. continuous number of) possible choices of $\mathbf{A}$ and A_0 is made leading to the same constant $\mathbf{E}$ and $\mathbf{B} = 0$. Of course the gauge invariance will assure that all physical quantities are independent of the gauge choice but in actual calculations it may require some efforts to see the connections.

5.1 Static gauge

We will explore in some detail two gauge choices, the simplest and most familiar gauge $\mathbf{A} = 0$, $A_0 = -\mathbf{E} \cdot \mathbf{r}$ and another, time-dependent gauge $\mathbf{A} = -\mathbf{E}t$, $A_0 = 0$. In the former case the time and the coordinate variables are separable in the Schrödinger equation

$$i\hbar \frac{\partial \psi(\mathbf{r}, t)}{\partial t} = \left(-\frac{\hbar^2}{2m} \nabla^2 - q\mathbf{E} \cdot \mathbf{r} \right) \psi(\mathbf{r}, t) \ , \tag{27}$$

and moreover also separable are the coordinates parallel and perpendicular to $\mathbf{E}$. Choosing the x axis parallel to $\mathbf{E}$ and denoting by subscript $\perp$ vectors which are perpendicular to $\mathbf{E}$ one can write the stationary solution as

$$\psi(\mathbf{r}, t) = \phi_\varepsilon(x) \exp\left(i\mathbf{k}_\perp \cdot \mathbf{r}_\perp\right) \exp\left[-\frac{it}{\hbar} \left(\varepsilon + \frac{\hbar^2 \mathbf{k}_\perp^2}{2m} \right) \right] \ , \tag{28}$$

where ε and $\phi_\varepsilon(x)$ are the eigenenergies and the corresponding eigenfunctions of the motion parallel to x. They satisfy the one dimensional Schrödinger equation

$$\left(-\frac{\hbar^2}{2m} \frac{d^2}{dx^2} - Fx \right) \phi_\varepsilon = \varepsilon \phi_\varepsilon \tag{29}$$

where we denoted $F = q \mid \mathbf{E} \mid$.

In the equation for $\phi_\varepsilon(x)$ the behavior of the potential $-Fx$ at infinite values of x is such that the energy levels ε form a continuous spectrum of values from $-\infty$ to $+\infty$. They should correspond to motion which is bounded from $x = -\infty$ but unbounded in the direction $x \to +\infty$. The wave functions must vanish in the region of large and negative x and therefore the energy levels are non degenerate. Indeed if there were two solutions $\phi_1(x)$ and $\phi_2(x)$ for the same ε then

$$\frac{1}{\phi_1} \frac{d^2 \phi_1}{dx^2} = \frac{2m}{\hbar^2} (\varepsilon + Fx) = \frac{1}{\phi_2} \frac{d^2 \phi_2}{dx^2} \tag{30}$$

so that the Wronskian $w = \phi_1(d\phi_2/dx) - \phi_2(d\phi_1/dx) = const.$ The condition that wave functions vanish at $x = -\infty$ means that $w = 0$ leading to $\phi_1 = const\ \phi_2$ i.e. the two solutions would in fact coincide.

5.2 Linear potential - the Airy function

The simplest way to solve the equation for ϕ_ε is to consider it in the momentum representation. Inserting the expansion

$$\phi_\varepsilon(x) = \int_{-\infty}^{\infty} \frac{dp}{\sqrt{2\pi\hbar}}\, a_\varepsilon(p)\, e^{ipx/\hbar} \tag{31}$$

in the equation for $\phi_\varepsilon(x)$ we easily obtain

$$\left(\frac{p^2}{2m} - i\hbar F\frac{\partial}{\partial p}\right) a_\varepsilon(p) = \varepsilon a_\varepsilon(p)\ . \tag{32}$$

Integrating this first order equation we find

$$a_\varepsilon(p) = const\ \ \exp\left[\frac{i}{\hbar F}\left(\varepsilon p - \frac{p^3}{6m}\right)\right]\ . \tag{33}$$

The constant in front of this expression must be determined by normalization. Choosing e.g., to normalize $a_\varepsilon(p)$ on the delta function in ε

$$\int_{-\infty}^{\infty} dp\ a^*_{\varepsilon'}(p)\ a_\varepsilon(p) = |C|^2 \int_{-\infty}^{\infty} dp\ \exp\left[\frac{i}{\hbar F}\left(\varepsilon - \varepsilon'\right) p\right] = \delta(\varepsilon - \varepsilon') \tag{34}$$

we obtain $C = 1/\sqrt{2\pi\hbar F}$. The wave functions in the position representation are

$$\phi_\varepsilon(x) = \int_{-\infty}^{\infty} \frac{dp}{2\pi\hbar\sqrt{F}}\ \ \exp\frac{i}{\hbar}\left[px + \frac{1}{F}\left(\varepsilon p - \frac{p^3}{6m}\right)\right] = \frac{\alpha}{\pi\sqrt{F}}\ \ Ai\left[-\alpha(x + \varepsilon/F)\right]\ , \tag{35}$$

where we denoted $\alpha = (2mF/\hbar^2)^{1/3}$ and introduced the notation

$$Ai(\xi) = \frac{1}{\sqrt{4\pi}} \int_{-\infty}^{\infty} du \exp\left[i\left(\frac{u^3}{3} + \xi u\right)\right]\ . \tag{36}$$

The function $Ai(\xi)$ defined by this integral is called the Airy function. We will explore some of its properties below, cf., also Ref. [11]. The graph of the Airy function is shown in Fig. 1.

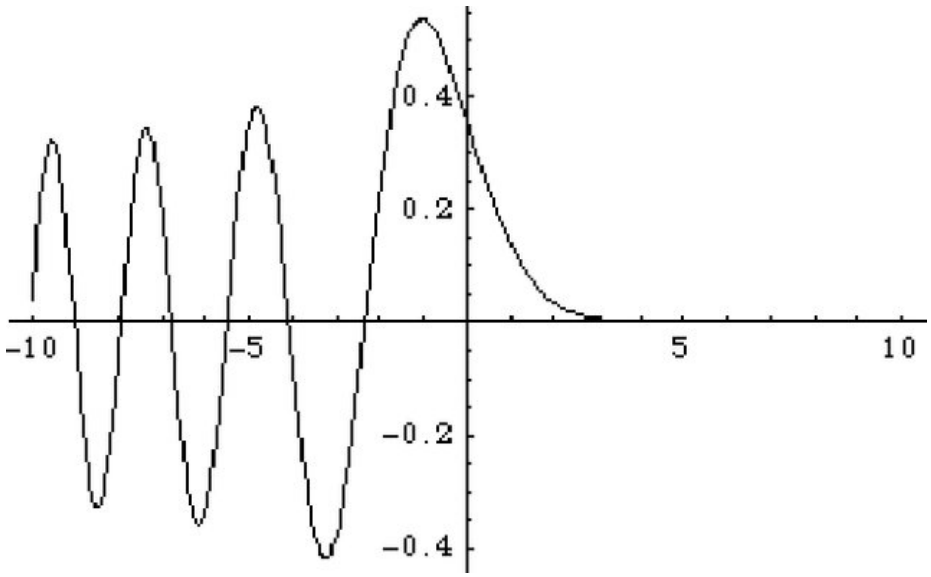

Figure 1: Airy function $Ai(x)$

Eq. (35) together with (28) furnish the general solution of the Schrödinger equation (27).

It is instructive to explore the asymptotic behavior of the wave function $\phi_\varepsilon(x)$ for $x \to \pm\infty$. This can be found by using the saddle point approximation

cf., https://atmos.washington.edu/ breth/classes/AM568/lect/lect22.pdf

in order to evaluate the integral in (35). Differentiating the exponent in the integrand we obtain that the stationary value p_0 of p must satisfy

$$\varepsilon = \frac{p_0^2}{2m} - Fx \; . \tag{37}$$

This is the classical – energy momentum relation in the potential $-Fx$. It is an example of a typical "cleverness" of the saddle point method – when a phase of a rapidly oscillating integral depends on external physical parameters (the coordinate x in (35)) the saddle point condition frequently has a transparent physically significance. The difference with the classical physics is that p_0 does not have to be real. Only for $x > -\varepsilon/F$, i.e. in the region where the classical motion is allowed, p_0 is real, but it is pure imaginary in the classically forbidden region $x < -\varepsilon/F$. In both cases there are two solutions corresponding to the two signs in the square root

$$p_0(x) = \sqrt{2m(\varepsilon + Fx)} \; . \tag{38}$$

According to the rules of the saddle point approximation both saddle point solutions should be retained in the real case while only the saddle point with decaying exponential should be admitted in the imaginary case. We thus find for (35)

$$\begin{aligned}\phi_\varepsilon(x) &\approx \sqrt{\frac{2m}{\pi\hbar p_0}}\cos\left(\frac{{p_0}^3(x)}{3\hbar Fm}-\frac{\pi}{4}\right) \ , \quad \text{for } x\to\infty \\ \phi_\varepsilon(x) &\approx \frac{1}{2}\sqrt{\frac{2m}{\pi\hbar|p_0|}}\exp\left(-\frac{|p_0(x)|^3}{3\hbar Fm}\right) \ , \ \text{for } x\to-\infty \ . \end{aligned} \tag{39}$$

We will see in the section devoted to the semiclassical limit that these expressions correspond to the semiclassical approximation for wave functions. As required the wave function decays exponentially in the classically forbidden region $E<-Fx$. In the classically allowed region the positive and negative momenta p_0 with equal amplitudes coexist for a stationary quantum mechanical state producing the interference cosine with the argument which can be written as

$$\frac{p_0^3(x)}{3\hbar Fm}\equiv\frac{1}{\hbar}W(x)=\frac{1}{\hbar}\int_{-E/F}^{x}p_0(x')\,dx' \tag{40}$$

in terms of the classical action W. The classical momentum determines the *local wave length* $\lambda=2\pi\hbar(dW/dx)^{-1}$ which decreases with increasing x in accordance with the uniform classical acceleration in the direction of the field and the de Broglie relation.

5.3 Time dependent gauge

Let us now examine how this problem looks in another gauge $\mathbf{A}=-\mathbf{E}t$, $A_0=0$. We use the gauge transformation (16) with $\chi=-\mathbf{E}\cdot\mathbf{r}t$ in the equation (27) and obtain

$$\begin{aligned}\psi(\mathbf{r},\mathrm{t}) &= \exp(iq\mathbf{E}\cdot\mathbf{r}t/\hbar)\psi'(\mathbf{r},\mathrm{t}) \ , \\ i\hbar\frac{\partial\psi'(\mathbf{r},\mathrm{t})}{\partial t} &= \frac{1}{2m}(-i\hbar\nabla+q\mathbf{E}t)^2\psi'(\mathbf{r},\mathrm{t}) \ , \end{aligned} \tag{41}$$

where we denoted by $\psi'(\mathbf{r},\mathrm{t})$ the transformed wave function.

A simple solution of this equation is a plane wave and we obtain for $\psi'(\mathbf{r},\mathrm{t})$

$$\psi'_{\mathbf{k}}(\mathbf{r},\mathrm{t})=A_{\mathbf{k}}(t)\exp[i\mathbf{k}\cdot\mathbf{r}] \tag{42}$$

with the time dependent amplitude $A_{\mathbf{k}}(t)$ satisfying

$$i\hbar\frac{dA_{\mathbf{k}}(t)}{dt}=\frac{1}{2m}(\hbar\mathbf{k}+q\mathbf{E}t)^2A_{\mathbf{k}}(t) \ . \tag{43}$$

Integrating we find

$$A_{\mathbf{k}}(t) = C_0 \exp\left\{ -\frac{i}{\hbar}\left[\frac{\hbar^2 \mathbf{k}_\perp^2}{2m}\, t \;\; + \;\; \frac{1}{6mF}(p+Ft)^3 \right] \right\} \;, \tag{44}$$

where C_0 is an arbitrary constant, $p = \hbar k_x$ and we used the same notation for $\mathbf{k}_\perp$ and F as in the previous section.

We note that with this solution the wave function before the gauge transformation (41) is

$$\psi_{\mathbf{k}}(\mathbf{r},\mathrm{t}) = A_{\mathbf{k}}(t) \exp[i(\mathbf{k} + q\mathbf{E}t/\hbar)\cdot \mathbf{r}] \tag{45}$$

which of course is a solution of the Schrödinger equation (27) in the static gauge. The set of these time dependent solutions with different $\mathbf{k}$'s is identical to the set (28) of stationary solutions as far as the motion in $\mathbf{r}_\perp$ is concerned. However in the direction of the field the sets look quite different and the point to note here is that in different gauges the same problem may have a very different appearance.

Of course mathematically both sets are equivalent and one can easily show that each can be expressed as a linear combination of the other. The time dependent solution is closer to the classical intuition of the accelerated motion under a constant force.

It is instructive in this simple problem to compare the calculations of the currents for two solutions $\psi_{\mathbf{k}}(\mathbf{r},\mathrm{t})$, Eq. (45), and the corresponding transformed one $\psi'_{\mathbf{k}}(\mathbf{r},\mathrm{t})$, Eq. (42). One will get different results with the two solutions

$$q\frac{\hbar\mathbf{k} + q\mathbf{E}t}{m}|C_0|^2 \quad \text{and} \quad q\frac{\hbar\mathbf{k}}{m}|C_0|^2$$

for the convective part of the current given by the fist part of Eq.(26). But this difference is "counterbalanced" by the different second diamagnetic term $-(q^2/m)\mathbf{A}|\psi|^2$ in the current expression. It is zero in the static gauge but is

$$q^2\frac{\mathbf{E}t}{m}|C_0|^2$$

in the time dependent gauge. The end results is of course the same expression as it must be for the gauge invariant quantity.

5.4 Translations in uniform E. Symmetries in the presence of gauge fields

Physics in a constant electric field must be invariant under translations of coordinates $\mathbf{r} \to \mathbf{r} + \mathbf{a}$ with a constant vector $\mathbf{a}$. Applying this transformation in the Schrödinger

equation (27) one at first finds that it is not invariant – the term $-\mathbf{E}\cdot\mathbf{a}$ is added to the Hamiltonian. This term however can be removed if one simultaneously performs a gauge transformation of the wave function

$$\psi(\mathbf{r},\mathrm{t}) \to \exp\left(-iq\mathbf{E}\cdot\mathbf{a}t/\hbar\right)\psi(\mathbf{r}+\mathbf{a},t) = \exp\left[i\mathbf{a}\cdot(-i\hbar\nabla - q\mathbf{E}t)/\hbar\right]\psi(\mathbf{r},\mathrm{t}) \qquad (46)$$

The Schrödinger equation is invariant under this combined transformation which must be therefore adopted as *the definition* of the translation in the present case. One can call it "electric translation" in analogy with the modified "magnetic translations" in a uniform magnetic field, cf. Section 6.3 below.

This feature of modification of the standard symmetry transformations by additional gauge transformations is quite typical for theories with gauge fields. It accounts for the fact that changing the coordinate system may also effect the gauge choice and care must be taken to return to the original gauge. The generator of the infinitesimal translations in the expression above is obviously

$$\mathbf{g}_{op} = \mathbf{p}_{op} - q\mathbf{E}t\ . \qquad (47)$$

The symmetry means that it must be conserved and indeed one finds that

$$d\mathbf{g}_{op}/dt = \partial\mathbf{g}_{op}/\partial t + (i/\hbar)[H\ ,\ \mathbf{g}_{op}] = 0$$

with H as it appears in the right hand side of (27).

We will find below two other examples of the gauge field modifications of the symmetries – translational and rotational invariance in a uniform magnetic field (Section 6.3) and rotational invariance in the field of a magnetic monopole (Section 11.3).

6 Motion in a Uniform Magnetic Field

We will now consider the quantum mechanical motion of a charged particle in a uniform external magnetic field $\mathbf{B}$ which is constant in magnitude and direction over the entire space. For convenience we present the discussion for electrons, i.e. we take the value of the charge

$$q = -e\ .$$

6.1 Classical motion. The guiding centers

It is instructive to recall first the classical solutions of the problem. The classical equation of motion is

$$m\,d\mathbf{v}/dt = -e(\mathbf{v}\times\mathbf{B})\ .$$

Let us choose the direction of the z axis parallel to $\mathbf{B}$. Then the motion along z is free,

$$m\, dv_z/dt = 0 \ , \ \ z = v_z t + z_0$$

with constant v_z and z_0 determined by initial conditions.

The equations for the x and y components are

$$\begin{aligned} m\frac{dv_x}{dt} &= -eBv_y \ , \\ m\frac{dv_y}{dt} &= eBv_x \ . \end{aligned} \tag{48}$$

An important observation to be made here is that these Newton equations for the velocities of motion in a plane perpendicular to $\mathbf{B}$ have the same formal appearance as the Hamilton equations of a *one dimensional oscillator* with v_x and v_y formally proportional to the respective coordinate and momentum of the oscillator. The solution of these equations is "harmonic motion" in the "velocity space" with the frequency

$$\omega_c = eB/m$$

called the cyclotron frequency and

$$v_x = v\cos(\omega_c t + \alpha) \ \ , \ \ v_y = v\sin(\omega_c t + \alpha) \ \ , \tag{49}$$

so that the trajectory in the (x,y) plane is

$$\begin{aligned} x &= \frac{v}{\omega_c}\sin(\omega_c t + \alpha) + x_0 = \frac{v_y}{\omega_c} + x_0 \ \ , \\ y &= -\frac{v}{\omega_c}\cos(\omega_c t + \alpha) + y_0 = -\frac{v_x}{\omega_c} + y_0 \end{aligned} \tag{50}$$

Here v, α, x_0 and y_0 are constants of the motion the values of which are fixed by the initial conditions $x(t_0), y(t_0), v_x(t_0), v_y(t_0)$ at some initial time t_0.

The physical meaning of these constants is the following. The above solution describes a circle with the radius v/ω_c. The position of the centre of the circle is given by the coordinates x_0 and y_0 which are therefore conventionally called the coordinates of the *guiding center*, cf., Fig. 2 below. The value of v also determines the energy

$$m(v_x^2 + v_y^2)/2 = mv^2/2$$

of the motion in the (x, y) plane[5]. This energy is independent of where on the x, y plane the orbit is situated, i.e. is independent of the values of x_0 and y_0 .

[5]The conservation of this quantity is trivially "discovered" by multiplying the two equations (48) respectively by v_x and v_y and adding

Using the terminology of quantum mechanics we can say that the above circular motion is degenerate - all circles with the same radius v/ω_c have the same energy. This degeneracy is characterized by different values of the guiding centre coordinates x_0 and y_0 so one can say that the classical motion is ∞^2 degenerate. As we will see in the next section in quantum mechanics the motion is "only" ∞ degenerate. It is important to observe that the expressions of the guiding centre coordinates as given by resolving (50)

$$x_0 = x - \frac{v_y}{\omega_c} \ , \ \ y_0 = y + \frac{v_x}{\omega_c} \tag{51}$$

are constants of the motion

$$\frac{dx_0}{dt} = \frac{dy_0}{dt} = 0$$

This fact, which is trivial in classical mechanics will play a very important role in the quantum mechanical treatment of the problem.

6.2 Landau levels

The quantum mechanics of this problem was first worked out by Landau and the corresponding solution is known as Landau levels.

6.2.1 The eigenenergies

The quantum Hamiltonian of a particle without spin in this case is

$$H_{op} = \frac{1}{2m}\left(-i\hbar\nabla + eA(\mathbf{r})\right)^2 = \frac{m\mathbf{v}_{op}^2}{2} \tag{52}$$

where the vector potential must be chosen such that $\mathbf{B} = \nabla \times \mathbf{A}$ is a constant vector parallel to z. With simple choices of $\mathbf{A}$ it is possible to find explicit solutions of the corresponding Schrödinger equation as we will discuss in detail below. At the moment however we prefer to proceed in a more general manner and show that many features of the solution can be anticipated on the basis of simple considerations which are useful to follow in order to gain a better understanding of the physics of the problem.

We start by considering the commutators of the components

$$\hat{v}_i = (-i\hbar\nabla_i + eA_i)/m$$

of the velocity operators which enter the Hamiltonian (52). They are easily calculated,

$$[\hat{v}_j, \hat{v}_k] = -\frac{ie\hbar}{m^2}\left\{[\nabla_j, A_k] + [A_j, \nabla_k]\right\} = -\frac{ie\hbar}{m^2}\left\{\frac{\partial A_k}{\partial x_j} - \frac{\partial A_j}{\partial x_k}\right\} = -i\frac{e\hbar}{m^2}\epsilon_{jkl}B_l \ . \tag{53}$$

These non vanishing commutators show that for a general magnetic field $\mathbf{B}(\mathbf{r})$ one can not have definite values simultaneously for all 3 components of the velocity.

In our particular case of a constant $\mathbf{B}$ along the z axis only the commutator $[\hat{v}_x, \hat{v}_y]$ is not zero. This means that $\hat{v}_z$ commutes with H_{op}, Eq. (52). Since moreover one can choose $A_z = 0$ and A_x and A_y to be functions of only (x, y) one has

$$H_{op} = \frac{m}{2}\left(\hat{v}_x^2 + \hat{v}_y^2\right) + \frac{\hat{p}_z^2}{2m} \quad , \quad \hat{p}_z = -i\hbar\frac{\partial}{\partial z}$$

so that the parts of H_{op} depending on (x, y) and on z are separable. The z–dependent part of the wave function must be a plane wave $\exp(ik_z z)$ with $k_z = mv_z/\hbar$ describing quantum free motion in accordance with the classical case.

The part of H_{op} describing the motion in the (x, y) plane,

$$h_{op} \equiv \frac{m}{2}\left(\hat{v}_x^2 + \hat{v}_y^2\right) \tag{54}$$

is proportional to the sum of squares of the operators $\hat{v}_x$ and $\hat{v}_y$ with a *constant commutator*

$$[\hat{v}_x\,,\,\hat{v}_y] = -i(e\hbar B/m^2)\,.$$

This suggests to define rescaled variables

$$\hat{p}_\xi = m\hat{v}_x \quad , \quad \hat{\xi} = (m/eB)\hat{v}_y$$

with the canonical commutator

$$[\hat{p}_\xi\,,\,\hat{\xi}] = -i\hbar$$

in terms of which the operator h_{op} takes the form

$$h_{op} = \frac{\hat{p}_\xi^2}{2m} + \frac{(eB)^2}{2m}\hat{\xi}^2 = \frac{\hat{p}_\xi^2}{2m} + \frac{m\omega_c^2}{2}\hat{\xi}^2$$

of the Hamiltonian of a one dimensional oscillator with mass m and frequency ω_c in accordance with the character of the corresponding classical motion. The spectrum of the oscillator is well known and adding it to the free motion eigenvalues of the $\hat{p}_z^2/2m$ term we obtain the eigenvalues of H_{op} as

$$E(n, k_z) = \hbar\omega_c\left(n + \frac{1}{2}\right) + \frac{\hbar^2 k_z^2}{2m}\,, \quad n = 0, 1, 2, ... \tag{55}$$

We have succeeded to obtain the eigenvalues of the Hamiltonian (52) on the basis of the commutation relations without solving the Schrödinger equation. There however remains a problem. The eigenvalues depend only on two quantum numbers n and k_z whereas dealing with three degrees of freedom one must find three quantum numbers which characterise the eigenfunctions of H_{op}.

6.2.2 Degeneracy of the Landau levels. Quantum guiding centers

The independence of $E(n, k_z)$ on the third quantum number suggests that the energy levels of the problem are degenerate and we will presently determine the reason and the nature of this degeneracy. For this purpose we will use the insights from the classical solutions and consider the quantum mechanical operators corresponding to the guiding center coordinates, Eq. (51),

$$\hat{x}_0 = x - \frac{\hat{v}_y}{\omega_c} \quad , \quad \hat{y}_0 = y + \frac{\hat{v}_x}{\omega_c} \, . \tag{56}$$

We easily find that they both commute with all the components of the velocity operators,

$$[\hat{x}_0, \hat{v}_i] = [\hat{y}_0, \hat{v}_i] = 0 \tag{57}$$

Indeed, e.g.

$$[\hat{x}_0, \hat{v}_x] = [x, \hat{v}_x] - \frac{1}{\omega_c}[\hat{v}_y, \hat{v_x}] = \frac{1}{m}[x, \hat{p}_x] - i\frac{\hbar}{m} = 0 \;\; ; \;\; [\hat{x}_0, \hat{v}_y] = [x, \hat{v}_y] = 0 \;\; , \;\; \text{etc}$$

Therefore $\hat{x}_0$ and $\hat{y}_0$ commute with H_{op}, i.e. are conserved as in the classical case. As in the classical treatment the energy of the motion is independent of these quantities. However we find that their commutator is not zero. Indeed using (57)

$$[\hat{x}_0, \hat{y}_0] = [\hat{x}_0, y] = -[\hat{v}_y, y]/\omega_c = -[\hat{p}_y, y]/m\omega_c = i(\hbar/m\omega_c) \tag{58}$$

This relation is commonly written as

$$[\hat{x}_0, \hat{y}_0] = i\ell^2 \tag{59}$$

where the constant

$$\ell = (\hbar/eB)^{1/2}$$

is called the magnetic length.

The non vanishing commutator between $\hat{x}_0$ and $\hat{y}_0$ means that they can not both have simultaneously definite values and moreover the constant value of the commutator shows that like the velocity operators above, their properties are similar to a canonical coordinate–momentum pair. Only one of the two can be specified and since it is conserved its eigenvalues should provide the missing quantum number which we are looking for in order to characterize the degenerate eigenfunctions belonging to the same eigenenergy $E(n, k_z)$. In fact the existence of the pair of *non commuting conserved* operators is the cause of the degeneracy of $E(n, k_z)$. If we choose the states

of the system to be eigenfunctions of, say, $\hat{x}_0$ operator, acting on one of them with $\hat{y}_0$ will produce a different state with the same energy. As we will show below there is a deep relation between the properties of the operators $\hat{x}_0$ and $\hat{y}_0$ and the basic symmetry of the system – the translational invariance.

6.2.3 The eigenfunctions

From the commutation relations Eq. (59) it follows that for the eigenstates with definite x_0 the values of y_0 are completely undetermined so that the position of the center of the quantized cyclotron orbit will have equal probability to be found at any point along the line with the given x_0. To see this explicitly we now turn to the solutions of the Schrödinger equation which have definite values of x_0. We need to choose first the gauge for the vector potential $\mathbf{A}$. The explicit forms of $\hat{x}_0$,

$$\hat{x}_0 \equiv x - \frac{\hat{v}_y}{\omega_c} = x - (1/eB)\hat{p}_y - A_y/B$$

and of

$$m\hat{v}_x = \hat{p}_x + eA_x$$

suggest the following convenient choice

$$A_x = 0\ ,\ A_y = Bx\ ,\ A_z = 0 \quad \rightarrow \quad \mathbf{B} = (0,0,B) \tag{60}$$

for which

$$\hat{x}_0 = i\ell^2\partial/\partial y \quad , \quad m\hat{v}_x = -i\hbar\partial/\partial x$$

and the Hamiltonian

$$H_{op} = \frac{\hat{p}_x^2}{2m} + \frac{1}{2m}\left(\hat{p}_y + eBx\right)^2 + \frac{\hat{p}_z^2}{2m}\ . \tag{61}$$

The eigenfunctions of $\hat{x}_0$ and $\hat{p}_z$ have the form

$$\psi(\mathbf{r}) = const\ \phi(x)\ e^{-ix_0y/\ell^2}e^{ik_zz} \quad , \quad const = \frac{1}{\sqrt{L_y}}\ \frac{1}{\sqrt{L_z}} \tag{62}$$

with yet undetermined $\phi(x)$. For convenience we have assumed that the motion in the y and the z directions is limited by large but finite intervals L_y and L_z with periodic boundary conditions.

Inserting in the Schrödinger equation $H_{op}\psi = E\psi$ and separating the variables we obtain

$$\left[\frac{\hat{p}_x^2}{2m} + \frac{m\omega_c^2}{2}\left(x - x_0\right)^2\right]\phi(x) = \varepsilon\phi(x)\ , \tag{63}$$

where we denoted $\varepsilon = E - (\hbar k_z)^2/2m$. This is the equation of a harmonic oscillator centered around the eigennvalue of x_0. As anticipated the eigenenergies are given by (55) and are independent of x_0. The eigenfunctions are

$$\phi_{n,x_0}(x) = \chi_n(x - x_0) \ , \tag{64}$$

where $\chi_n(x)$ are the normalized eigenfunctions of harmonic oscillator

$$\chi_n(x) = \left(\frac{1}{\pi\ell^2}\right)^{\frac{1}{4}} \frac{1}{\sqrt{2^n n!}} \exp\left[-x^2/\ell^2\ \right]\ H_n\left[x/\ell\ \right] \tag{65}$$

with $H_n(x) = (-1)^n\ e^{x^2}(d^n/dx^n)e^{-x^2}$ – the Hermite polynomials. The first few functions $\chi_n(x)$ are

$$\begin{aligned} \chi_0(x) &= \left(\frac{1}{\pi\ell^2}\right)^{\frac{1}{4}} \exp\left(-\frac{x^2}{2\ell^2}\right)\ ,\ \chi_1(x) = \left(\frac{1}{\pi\ell^2}\right)^{\frac{1}{4}} \frac{\sqrt{2}x}{\ell} \exp\left(-\frac{x^2}{2\ell^2}\right)\ , \\ \chi_2(x) &= \left(\frac{1}{4\pi\ell^2}\right)^{\frac{1}{4}} \left(\frac{x^2}{\ell^2} - 1\right) \exp\left(-\frac{x^2}{2\ell^2}\right)\ ,\ \text{etc}\ . \end{aligned} \tag{66}$$

Note that the usual quantum harmonic oscillator length scale $(\hbar/m\omega)^{1/2}$ controlling the relative extent of its wavefunctions is magnetic length

$$(\hbar/m\omega_c)^{1/2} = l$$

in this case.

Imposing periodic boundary conditions in the y direction we find that x_0 in Eq.(62) takes discrete values separated by distances $\Delta x_0 = 2\pi\ell^2/L_y$. We thus have one state per area $L_y\Delta x_0 = 2\pi\ell^2$ in the x-y plane. The dependence of the wave functions on y via the plane wave phase means that the probability to find a particle is independent of this coordinate. It also seems to suggest that like in the z - direction there is a free motion also in the y-direction. This however is not correct as it is based on the experience in situations in which there was no gauge field present. In this case the wavefunction's phase is gauge dependent so to evaluate what motions it describes one must form gauge invariant observables. We will do this below by calculating the current density components with physically interesting results.

In the x direction the state is centered around the value x_0. Its extension can be determined, using e.g., the equipartition property of the oscillator meaning that the average potential energy is one half of the total energy, $m\omega_c^2\langle(x - x_0)^2\rangle/2 =$

$\hbar\omega_c(n+1/2)/2$. This gives $\sqrt{\langle(x-x_0)^2\rangle} = \ell\sqrt{n+1/2}$. Each degenerate energy level can thus pictorially be viewed as a two dimensional plane filled with overlapping (for $L_y \gg 2\pi\ell/\sqrt{n+1/2}$) "strips" occupied by individual quantum states parallel to the y axis representing quantized cyclotron orbits uniformly "smeared" along every strip. This picture repeats itself for every n and k_z with the radius of the orbits, i.e. the thickness of the strips growing as $\ell\sqrt{n+1/2}$. The smearing of the orbits is the result of the Heisenberg–like uncertainty relation between the guiding center coordinates x_0 and y_0.

The degenerate energy levels which we have just described are called Landau levels. Choosing y_0 to have defined values will lead to the same picture of Landau levels but with the strips parallel to the x axis. It is amusing to consider what happens if more complicated functions of x_0 and y_0 are chosen to have defined values. Suppose we fix $x_0^2 + y_0^2$. Then the strips in the picture above will have the shape of concentric circles around the origin. Choosing fixed $x_0^2/a^2 + y_0^2/b^2$ with some constants a and b will lead to strips of elliptic shapes, while fixing the function $(x_0y_0 + y_0x_0)/2$ (symmetrized to make the corresponding operator hermitian) will result in a hyperbolic shape of the strips, etc.

In Fig.2 we illustrate some of these cases. Of course all the choices above are equivalent as long as the degeneracy remains but some may be singled out if a perturbation removing this degeneracy is added to the Hamiltonian.

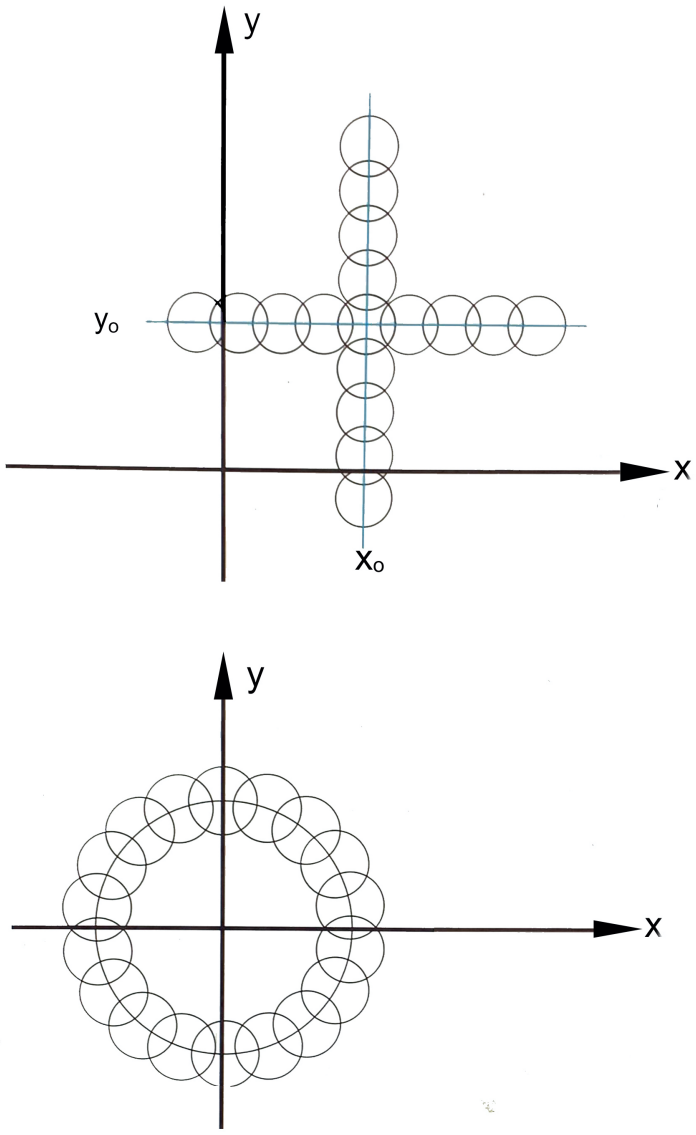

Figure 2: Schematic illustration of a single classical cyclotron orbit and how it gets "smeared" in quantum mechanical description. The upper figure shows the cases of x_0 fixed (an orbit is smeared in the y-direction) or y_0 fixed (orbit is smeared in the x-direction). In the lower figure $x_0^2 + y_0^2$ is fixed - an orbit is smeared along the corresponding circle

When choosing the eigenvalues of the operator $\hat{y}_0$ instead of $\hat{x}_0$ for the characterization of the degenerate wave functions it should be more convenient to choose the gauge $A_x = -By$, $A_y = A_z = 0$ in which $\hat{y}_0$ is just $-i\ell^2\partial/\partial x$. For a combination $\hat{x}_0^2 + \hat{y}_0^2$ the symmetric choice $A_x = -By/2$, $A_y = Bx/2$, $A_z = 0$ is the most appropriate. In the literature it may sometimes seem that the choice of the gauge determines which combination of $\hat{x}_0$ and $\hat{y}_0$ will be diagonal. Our remark here is meant to clarify the correct order of choices .

The square of the magnetic length ℓ appearing in the commutator of the guiding center coordinates plays the role of the "Planck constant" for these variables. Therefore the analogue of uncertainty relation $\Delta x_0 \Delta y_0 \geq \ell^2/2$ must hold. We also recall from the statistical physics that in the semiclassical picture quantum states "occupy" phase space volume $\Delta p \Delta q = 2\pi\hbar$. Here we may expect an analogous situation that a single state of a degenerate Landau level "occupies" an area in the plane of (x_0, y_0) which is $2\pi\ell^2$. And indeed we have seen this in the particular case of the solutions (62). The physical meaning of this minimal area is simple and profound – the magnetic flux through this area is ratio of universal world constants

$$B \cdot 2\pi\ell^2 = 2\pi B \frac{\hbar}{eB} = \frac{2\pi\hbar}{e} = \frac{h}{e}$$

Such magnetic flux has a special notation

$$\Phi_0 = \frac{2\pi\hbar}{e} = \frac{h}{e} \tag{67}$$

and a special name - magnetic flux quantum. We will meet this quantity a number of times in these notes (cf., below). Let us stress that its name doesn't mean that the magnetic flux in such problems is quantized. Rather, as we see here and will be seen below certain physical features get repeated with Φ_0 as a period.

As can be seen from the above discussion the density of single states in a degenerate Landau level is the inverse of $2\pi\ell^2$ which is independent of n and of the way we choose to classify the degeneracy. The mnemonic rule of "one state per one flux quantum" is something which is encountered in many quantum mechanical problems in the presence of magnetic field and is therefore well worth remembering.

6.2.4 Currents and edge currents

Individual states in a Landau level carry a non vanishing *current density*. Apart from an obvious contribution from the free motion in the z–direction one also finds current distribution in the x-y plane. Qualitatively one expects that in this plane

the quantum mechanically smeared cyclotron orbits with one fixed guiding center coordinate should combine to give opposite currents parallel to and concentrated on the edges of the strip occupied by the state and have zero current on the midline of the strip.

We easily find for the states Eq. (62) using Eq. (26) for the current

$$j_x = 0 \ , \ j_y(x) = e\omega_c(x_0 - x)\rho(x) \ , \ j_z(x) = e(\hbar k_z/m)\rho(x) \tag{68}$$

where we denoted the particle density

$$\rho(x) = \phi^2(x)/L_y\, L_z$$

The appearance of the lengths L_y and L_z is related to the (standard) normalization of the wavefunction (62) to one particle.

The wave function $\phi^2(x)$ as given by any of the solutions Eq. (64) is concentrated in a symmetric strip around x_0 which means that the current density $j_y(x)$ has an antisymmetric profile with respect to $x = x_0$. Because of this antisymmetry the *total current*

$$I_y = \int j_y(x)\, dx \tag{69}$$

flowing in the y-directions, i.e. along the state ψ_{n,x_0} in the x-y plane is zero for these states.

If one adds a constant electric field parallel to the x–axis one can still find exact wave functions (cf., homework problems or tutorial). The current density profile of these wave functions will change from antisymmetric to asymmetric and the total current *in the y direction* will not be zero.

Another interesting non zero current carrying Landau states appear at the edges of the x-y plane. Let us assume that an additional potential $V(x)$ with the shape shown in Fig. 3 is added to the Eq. (63)

$$\left[\frac{\hat{p}_x^2}{2m} + \frac{m\omega_c^2}{2}\,(x - x_0)^2 + V(x)\right]\phi(x) = \varepsilon\phi(x) \ , \tag{70}$$

This potential simulates the edges of the sample in the x direction.

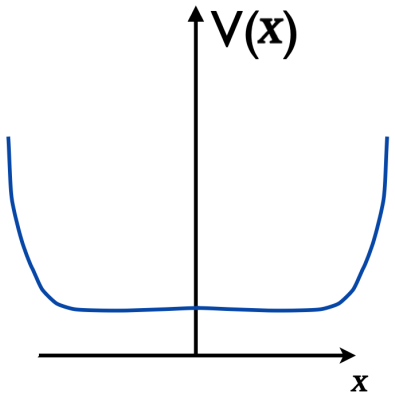

Figure 3: Potential simulating edges in the x-y plane

It is instructive to examine how the combined potential

$$U(x) = \frac{m\omega_c^2}{2}(x - x_0)^2 + V(x)$$

changes when plotted for different guiding center coordinate values x_0 relative to the positions of the potential walls representing the edges. The shape of $U(x)$ getting more narrow for the values of x_0 near and "inside" the edges indicates that eigenenergies ϵ_n will break the degeneracy of the Landau levels in such a way that they become rising functions $\epsilon_n(x_0)$ for such values of x_0. Recalling the width $\ell\sqrt{n+1/2}$ which the unperturbed Landau levels occupy we can approximate for low n values and the potential $V(x)$ slowly varying on the magnetic length ℓ scale as

$$\begin{aligned} V(x) &\approx V(x_0) \\ \epsilon_n(x_0) &\approx \hbar\omega_c(n + \frac{1}{2}) + V(x_0) \end{aligned} \tag{71}$$

In this approximation the modified Landau levels similar to the unperturbed ones form an "equidistant ladder" with each step having the shape of $V(x_0)$.

The asymmetric shape of the combined $U(x)$ for x_0 near the edges means that the resulting eigenfunctions $\phi_{n,x_0}(x)$ will not depend on x_0 via $x - x_0$ as in Eq. (64) and will not have the harmonic oscillator symmetry around x_0 as in the unperturbed Landau states. This in turn means that the current density $j_y(x)$ along the edge will not have an antisymmetric profile with respect to $x = x_0$ and therefore the

total current flowing in the x-y plane for such states near the edges will not be zero. Such currents are called "edge currents". They correspond to the skipping classical orbits near potential walls, move in opposite directions on the opposite edges and play important role in explaining the Quantum Hall Effect, cf., Ref. [14].

Let us express the energy of a given state ϕ_{n,x_0} using

$$h_{op}(x_0)\phi_{n,x_0} = \epsilon_n(x_0)\phi_{n,x_0} \quad \rightarrow \quad \epsilon_n(x_0) = \langle\phi_{n,x_0}|h_{op}(x_0)|\phi_{n,x_0}\rangle$$

with $h_{op}(x_0)$ denoting the Hamiltonian operator in the left hand side of Eq. (70). Using the Feynman-Hellman theorem[6] one obtains

$$\begin{aligned}\frac{\partial\epsilon_n(x_0)}{\partial x_0} &= \langle\phi_{n,x_0}|\frac{\partial h_{op}}{\partial x_0}|\phi_{n,x_0}\rangle = \\ &= m\omega_c^2\langle\phi_{n,x_0}|(x-x_0)|\phi_{n,x_0}\rangle = m\omega_c^2\int dx(x-x_0)\,\phi^2_{n,x_0}(x)\end{aligned} \tag{72}$$

Comparing with the expression for the current density $j_y(x)$ in Eq. (68) and ignoring for convenience the z direction we find the relation

$$\frac{\partial\epsilon_n(x_0)}{\partial x_0} = \frac{m\omega_c L_y}{e}\; I_y(n,x_0) \tag{73}$$

where $I_y(n,x_0)$ is the total current of a single particle in the ψ_{n,x_0} state. Referring to Eq. (71) with $V(x)$ as shown in Fig. 3 one sees clearly where the edge currents are expected, their magnitude and direction.

[6]The theorem relates the derivative of the eigenenergy with respect to a parameter to the expectation value of the derivative of the Hamiltonian with respect to that parameter. The proof is straightforward

$$\begin{aligned}\frac{\partial\epsilon_n(x_0)}{\partial x_0} &= \frac{\partial}{\partial x_0}\langle\phi_{n,x_0}|h_{op}(x_0)|\phi_{n,x_0}\rangle = \\ &= \langle\frac{\partial\phi_{n,x_0}}{\partial x_0}|h_{op}(x_0)|\phi_{n,x_0}\rangle + \langle\phi_{n,x_0}|h_{op}(x_0)|\frac{\partial\phi_{n,x_0}}{\partial x_0}\rangle + \langle\phi_{n,x_0}|\frac{\partial h_{op}}{\partial x_0}|\phi_{n,x_0}\rangle = \\ &= \epsilon_n(x_0)\left[\langle\frac{\partial\phi_{n,x_0}}{\partial x_0}|\phi_{n,x_0}\rangle + \langle\phi_{n,x_0}|\frac{\partial\phi_{n,x_0}}{\partial x_0}\rangle\right] + \langle\phi_{n,x_0}|\frac{\partial h_{op}}{\partial x_0}|\phi_{n,x_0}\rangle = \langle\phi_{n,x_0}|\frac{\partial h_{op}}{\partial x_0}|\phi_{n,x_0}\rangle\end{aligned}$$

where it was used that

$$\frac{\partial}{\partial x_0}\langle\phi_{n,x_0}|\phi_{n,x_0}\rangle = 0 = \langle\frac{\partial\phi_{n,x_0}}{\partial x_0}|\phi_{n,x_0}\rangle + \langle\phi_{n,x_0}|\frac{\partial\phi_{n,x_0}}{\partial x_0}\rangle$$

6.3 Degeneracy of Landau levels and space symmetries

Conservation laws are always results of symmetries and the existence of the conserved operators $\hat{x}_0$, $\hat{y}_0$ and $\hat{v}_z$ is not an exception. They are related to the basic symmetry of the motion in a uniform field – invariance under translations. This invariance is however not explicit in the Hamiltonian (52) which changes under the translation $\mathbf{r}$ to $\mathbf{r}+\mathbf{a}$ with an arbitrary constant vector $\mathbf{a}$. We have already encountered a similar phenomenon in the simpler case of a uniform electric field. Also here the the Hamiltonian (52) remains invariant if simultaneously with the proper translation one performs a suitably chosen gauge transformation. The conserved quantities should be the appropriate generators of these combined transformations.

To see this in detail we observe that after a proper translation the Schrödinger equation with the Hamiltonian (52) has the same form but with the different vector potential $\mathbf{A}(\mathbf{r}+\mathbf{a})$. *For a constant* $\mathbf{B}$ however the difference $\mathbf{A}(\mathbf{r}+\mathbf{a})-\mathbf{A}(\mathbf{r})$ is a gauge transformation, i.e. it is a gradient of a scalar function. It will be sufficient to show this for an infinitesimal $\mathbf{a}$ for which we have $\mathbf{A}(\mathbf{r}+\mathbf{a}) \approx \mathbf{A}(\mathbf{r}) + (\mathbf{a}\cdot\nabla)\mathbf{A}(\mathbf{r})$. The last term is

$$a_i \partial_i A_j = a_i(\partial_i A_j - \partial_j A_i) + a_i \partial_j A_i = a_i(\epsilon_{ijk} B_k + \partial_j A_i)$$

and for a constant $\mathbf{B}$ it is indeed a gradient

$$\partial_j \alpha \quad \text{with} \quad \alpha = a_i(\epsilon_{ijk} x_j B_k + A_i) = \mathbf{a}\cdot[\mathbf{r}\times\mathbf{B} + \mathbf{A}(\mathbf{r})] \tag{74}$$

It can be removed from H_{op} by a gauge transformation of the wave function in addition to the proper translation. The symmetry transformation is therefore

$$\psi(\mathbf{r}) \to \left\{1 + \frac{ie}{\hbar}\mathbf{a}\cdot[\mathbf{r}\times\mathbf{B} + \mathbf{A}(\mathbf{r})]\right\}(1 + i\mathbf{a}\cdot\mathbf{p}_{op}/\hbar)\psi(\mathbf{r}) \quad (\text{infinitesimal} \quad \mathbf{a})\,, \tag{75}$$

where we have used the proper translation operator $\exp(i\mathbf{a}\cdot\mathbf{p}_{op}/\hbar)$ for infinitesimal $\mathbf{a}$ to write $\psi(\mathbf{r}+\mathbf{a})$ in terms of $\psi(\mathbf{r})$.

The combined transformation (75) is what should be called translation in the presence of a magnetic field (the term "magnetic translation" is sometimes used). The generators of this transformation are read off the linear term in $\mathbf{a}$ found after multiplying the brackets in Eq.(75). They are

$$\mathbf{p}_{op} + e[\mathbf{A}(\mathbf{r}) + \mathbf{r}\times\mathbf{B}] = m\mathbf{v}_{op} + e\mathbf{r}\times\mathbf{B} \tag{76}$$

For $\mathbf{B}||\mathbf{e}_z$ and translations along the z axis this is just mv_z whereas for the translations along the x and y axes we obtain respectively $eB\,\hat{y}_0$ and $-eB\,\hat{x}_0$ in terms of the operators of the guiding center coordinates.

It should now become intuitively clear why these operators do not commute. We expect that the result of translating the wave function parallel to x and then parallel to y should not be the same as translating it in the opposite order. The difference should be related to the Aharonov-Bohm phase (see Section 7 below for its definition) induced by the flux of the magnetic field through the rectangle obtained in the course of these reversed order translations. Let us see how it happens. Transporting a wave function by an infinitesimal Δx followed by Δy and then by $-\Delta x$ and $-\Delta y$ respectively one indeed obtains keeping the terms up to a 2nd order in the translations Δx and Δy

$$
\begin{aligned}
&(1 - i\Delta y K_y - \frac{1}{2}(\Delta y)^2 K_y^2)(1 - i\Delta x K_x - \frac{1}{2}(\Delta x)^2 K_x^2) \times \\
&\times (1 + i\Delta y K_y - \frac{1}{2}(\Delta y)^2 K_y^2)(1 + i\Delta x K_x - \frac{1}{2}(\Delta x)^2 K_x^2)\psi(\mathbf{r}) = \\
&= (1 + \Delta x \Delta y [K_x\,,\, K_y])\psi(\mathbf{r}) = [1 + 2\pi i(\Delta\Phi/\Phi_0)]\,\psi(\mathbf{r})
\end{aligned} \tag{77}
$$

where we denoted by $\hbar K_x$ and $\hbar K_y$ the corresponding vector components of the operator of translations (76) and $\Delta\Phi = B\Delta x\Delta y$ – is the flux through the rectangle. Since the first non vanishing term in the expression above apart of unity was quadratic and proportional to $\Delta x\Delta y$ it was necessary to keep the quadratic terms in the operators of each translation.

A similar discussion concerning the generalization of transformations and their generators can be worked out for another symmetry of the problems – the rotational symmetry around the direction of the magnetic field $\mathbf{B}$. We will leave this for homework or tutorials.

7 The Aharonov - Bohm Effect

7.1 Local and non local gauge invariant quantities

We have emphasized in Section 3.3 that all observable quantities in a theory with a gauge field are gauge invariant. Perhaps the simplest such quantities are the electric and magnetic fields and the particle density $\rho(\mathbf{r},\mathrm{t}) = |\psi(\mathbf{r},t)|^2$. In the expression for the electric current density considered in the previous section we encountered another set involving the derivatives of ψ – the combinations $\psi^*(\mathbf{r},\mathrm{t})\mathbf{D}\psi(\mathbf{r},\mathrm{t})$ to which we can also add their time dependent partner $\psi^*(\mathbf{r},\mathrm{t})D_0\psi(\mathbf{r},\mathrm{t})$. These combinations are gauge invariant due to the simple transformation properties of the gauge covariant derivatives (17).

A distinct feature of all these invariants is that they depend on $\psi, \mathbf{A}$ and A_0 and their first derivatives at the *same space–time point*, i.e. they are *local*. Consider however a circulation integral $\oint \mathbf{A} \cdot d\mathbf{r}$ taken around some closed contour drawn in space. By the Stockes theorem this integral is equal to the flux of $\mathbf{B}$ through the contour and is therefore gauge invariant. This is an example of a *non local* gauge invariant quantity. In the following sections we will discuss situations in which the non trivial dependence on $\oint \mathbf{A} \cdot d\mathbf{r}$ leads to unexpected quantum mechanical effects which are collectively known as the Aharonov–Bohm effect, Ref. [7]. The sensitivity of the quantum theory to non local gauge invariants can be traced to essential non locality of the quantum mechanical description — eigenvalues and expectation values of various physical quantities such as energy, angular momentum, etc., depend on what happens with the wave function in the entire configuration space of the system.

Concluding this section we mention that in addition to the circulation of the vector potential another type of non local gauge invariants appears in certain physical applications. These are bi-local quantities of the type

$$\psi^*(\mathbf{r}') \exp[i(e/\hbar) \int_{\mathbf{r}}^{\mathbf{r}'} \mathbf{A}(\mathbf{r}'') \cdot d\mathbf{r}'']\psi(\mathbf{r}, t) \ .$$

Under gauge transformations the exponential and the wave functions produce phase factors which cancel each other. Quantities like this are often met in the field theoretical context and recently in certain many body problems.

7.2 Quantum mechanics "feels" non zero $\oint_C \mathbf{A} \cdot d\mathbf{r}$ even if $\mathbf{E} = \mathbf{B} = 0$ on and near the contour C

Let us consider a region of space in which local invariant quantities $\mathbf{E}$ and $\mathbf{B}$ are zero but in which contours can be found for which $\oint \mathbf{A} \cdot d\mathbf{r}$ does not vanish. A simple example is a region *outside* of a long thin tube with impenetrable walls and a non zero magnetic field concentrated inside and running parallel to the tube, Fig.4

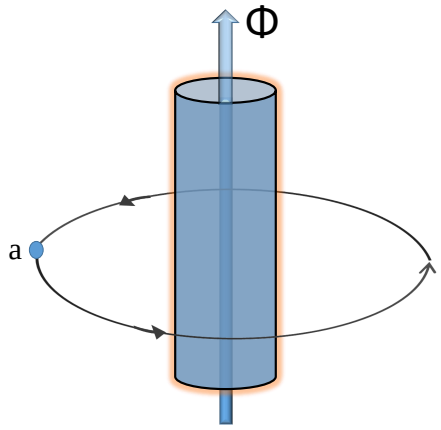

Figure 4: Example of the Aharonov-Bohm flux Φ in an impenetrable tube and a closed contour encircling it in the region with zero $\mathbf{E}$ and $\mathbf{B}$. The non zero circulation $\oint_C \mathbf{A} \cdot d\mathbf{r} = \Phi$ around such contours has no effect in classical description of charged particle motion (trajectories) in this region but produces observable effects in its quantum mechanics (wave functions)

Non zero circulation integrals $\oint \mathbf{A} \cdot d\mathbf{r}$ are obtained for the integration contours which wind around the tube. Since by assumption $\mathbf{B} = \nabla \times \mathbf{A} = 0$ outside the tube the details of a particular contour are of no importance except for the number of times it winds around the tube and the direction of this winding. Denoting this number by n one can write

$$\oint_C \mathbf{A} \cdot d\mathbf{r} = n\Phi \quad , \quad n = 0, \pm 1, ... \tag{78}$$

Here Φ denotes the magnitude of the total flux of the magnetic field in the tube. The circulation integrals outside the tube depend only on Φ and not to the details of the magnetic field distribution. One conventionally refers to such a tube as a solenoid and to such an isolated magnetic flux Φ as the Aharonov–Bohm flux (AB flux for brevity).

7.3 "Gauging out" the AB flux. Periodic dependence on its value

Classically the *free motion* of a particle in the outside region is not influenced by the presence of the field inside the tube. At first sight one may reach a similar conclusion in the quantum mechanical description. Indeed to write the Schrödinger equation one needs to determine first the electromagnetic potentials. Since $\mathbf{B} = \nabla \times \mathbf{A} = 0$ in the outside region one must have that $\mathbf{A}$ must be a gradient of some scalar,

$$\mathbf{A} = \nabla \xi(\mathbf{r}) \tag{79}$$

With such $\mathbf{A}$ (and $\mathbf{E} = 0$) it may appear that in the corresponding Schrödinger equation

$$i\hbar \frac{\partial \psi}{\partial t} = \frac{1}{2m} \left(-i\hbar\nabla + e\mathbf{A}(\mathbf{r})\right)^2 \psi$$

one could remove the $e\mathbf{A}$ term by a gauge transformation

$$\psi(\mathbf{r}, t) = \psi'(\mathbf{r}, t) \exp[-i\frac{e}{\hbar}\xi(\mathbf{r})]$$

with

$$\xi(\mathbf{r}) = \xi_0 + \int_{\mathbf{r}_0}^{\mathbf{r}} \mathbf{A}(\mathbf{r}') \cdot d\mathbf{r}' = \xi_0 + \int_{\mathbf{r}_0}^{\mathbf{r}} \nabla\xi(\mathbf{r}') \cdot d\mathbf{r}'$$

and ξ_0 some constant.

The problem however with this elimination of $\mathbf{A}$ from the Schrödinger equation is that in the presence of the AB flux Φ the scalar function $\xi(\mathbf{r})$ in Eq. (79) is not single valued. It is a *multivalued* function as can be seen in the following way. To have the required value of the AB flux the function $\xi(\mathbf{r})$ must change by Φ when "taken (followed) continuously" along a contour C around the solenoid in the positive direction

$$\Phi = \oint_C \mathbf{A} \cdot d\mathbf{r} = \oint_C \nabla\xi(\mathbf{r}) \cdot d\mathbf{r} = \int_{\mathbf{r}_i}^{\mathbf{r}_f} \nabla\xi \cdot d\mathbf{r} = \xi(\mathbf{r}_f) - \xi(\mathbf{r}_i) \quad \text{with} \quad \mathbf{r}_f = \mathbf{r}_i \ . \tag{80}$$

Thus at every $\mathbf{r}$ in the region outside the solenoid the function $\xi(\mathbf{r})$ has many (infinity) of values differing by $n\Phi$ with (positive or negative) integer n.

Given this the transformed $\psi'(\mathbf{r}, t)$,

$$\psi'(\mathbf{r}, t) = \psi(\mathbf{r}, t) \exp[i\frac{e}{\hbar}\xi(\mathbf{r})]$$

will also be multivalued - its phase will change by

$$\Delta\xi = \frac{e}{\hbar}\Phi = 2\pi\frac{\Phi}{\Phi_0} \tag{81}$$

when "taken continuously" around the solenoid.

To understand what the demand of such a particular non single valuedness of the wave function produces let us consider a specific example of the angular momentum. Assuming the z-axis along the solenoid and the z component $\hat{L}_z$ we have for its eigenfunctions

$$\hat{L}_z\psi(\phi) = \hbar\nu\psi(\phi) \quad \rightarrow \quad \psi(\phi) = const\, e^{i\nu\phi}$$

In the usual case, i.e. in the absence of the AB flux one applies the condition $\psi(\phi + 2n\pi) = \psi(\phi)$, i.e. the condition of single valuedness of $\psi(\phi)$ which leads to the usual integer quantization

$$\nu = M \ , \quad M = 0, \pm 1. \pm 2, ...$$

For the multivalued function condition Eq. (81) we have

$$\begin{aligned} e^{i\nu(\phi+2n\pi)} =& e^{i\nu\phi}e^{i2\pi n\Phi/\Phi_0} \rightarrow 2n\pi(\nu - \Phi/\Phi_0) = 2n\pi M \rightarrow \\ \rightarrow\ & \nu = M + \Phi/\Phi_0 \end{aligned} \tag{82}$$

This shows that despite our "gauging out" of the vector potential $\mathbf{A} = \nabla\xi$ its gauge invariant content, i.e. the AB flux Φ in Eq. (80), if not zero modifies the physics via the resulting multivalued wave function condition Eq. (81). We will see below that this modifications is (not surprisingly) identical to the straightforward solution with such a vector potential.

As an important additional observation we note that when $\Phi = n\Phi_0$ there is no effect! The transformed ψ' remains single valued and such AB flux is non observable "from outside". This observation is probably one of the advantages of the multivalued wave function formulation. It also means that the Aharonov-Bohm effects have periodic dependence on the magnitude Φ of the AB flux with the period of the flux quantum Φ_0. One can see this in the dependence of the eigenvalues ν on Φ, Eq. (82). They change with Φ from integer to integer with the period Φ_0. We will also see this periodicity in the examples considered in the next section and will provide a more general point of view in Section 7.5.

Another important observation is the following. The view of the Aharonov–Bohm effect as a modification of the condition that the wave function repeats itself as it is taken around a solenoid stresses that in order to "feel" this modification the wave function must extend all around the solenoid. Otherwise there will be no observable

consequences of the Aharonov–Bohm flux. Below we will consider an example of a ring pierced by the Aharonov–Bohm flux with a particle localized on a finite sector of the ring. There is no Aharonov–Bohm effect in this case.

7.4 Example of the AB flux

Assume that the solenoid with the AB flux Φ is placed along the z-axis. A possible simple choice for the vector potential outside such a solenoid is

$$\mathbf{A} = \nabla\xi(\mathbf{r}) \ , \quad \text{with} \ \ \xi(\mathbf{r}) = \frac{\Phi}{2\pi}\arctan(y/x) \equiv \frac{\Phi}{2\pi}\phi \tag{83}$$

where ϕ is the azimuthal angle. Recalling the expression of the gradient in cylindrical coordinates r, ϕ, z

$$\nabla = \mathbf{e}_r\frac{\partial}{\partial r} + \mathbf{e}_\phi\frac{1}{r}\frac{\partial}{\partial\phi} + \mathbf{e}_z\frac{\partial}{\partial z}$$

one finds

$$A_\phi = \frac{\Phi}{2\pi r} \ \ ; \ \ A_r = A_z = 0 \tag{84}$$

and therefore the circulation integral outside the solenoid along a circular contour in a plane perpendicular to the solenoid

$$\oint \mathbf{A}\cdot d\mathbf{r} = \int_0^{2\pi} A_\phi \, r d\phi = \Phi$$

Since $\nabla\times\mathbf{A} = \mathbf{B} = 0$ outside the solenoid one can deform the above circular contour without changing the integral as long as the new contour has "the same topology" - i.e. encircles the flux once in the same direction. One can also change the particular $\mathbf{A}$ in (83) by adding a *single valued function* to ξ without influencing $\mathbf{B} = 0$ or circulation integrals $\oint \mathbf{A}\cdot d\mathbf{r}$ outside the solenoid. We observe that the dependence of the outside vector potential on the magnetic field is via the flux Φ irrespective of a particular radial dependence of $\mathbf{B}$ inside the solenoid.

To have a convenient example of the AB flux one can think of $\mathbf{B} = B(r)\mathbf{e}_z$ with a constant B inside and zero outside. With this magnetic field one can write for all r's,

$$\begin{aligned} \mathbf{A} &= (Br/2)\,\mathbf{e}_\phi \qquad \text{inside the solenoid} \, , \\ \mathbf{A} &= (Br_0^2)/2r\ \mathbf{e}_\phi = \frac{\Phi}{2\pi r}\mathbf{e}_\phi \quad \text{outside the solenoid} \, , \end{aligned} \tag{85}$$

where r_0 is the radius of the solenoid. Since $\Phi = B(\pi r_0^2)$, this expression for the outside region is the same as (83).

7.4.1 The Hamiltonian and the spectrum

The Hamiltonian with the vector potential (84) outside the AB flux has a simple form in cylindrical coordinates. Using $\mathbf{p} = p_r\mathbf{e}_r + p_\phi\mathbf{e}_\phi + p_z\mathbf{e}_z$ and $\mathbf{A} = A_\phi\mathbf{e}_\phi$ we have

$$H = \frac{p_r^2}{2m} + \frac{1}{2m}\left(p_\phi + eA_\phi\right)^2 + \frac{p_z^2}{2m} + U(r)\ , \tag{86}$$

where we disregarded the spin degrees of freedom and added U(r) - the potential which should account for the impenetrable walls of the solenoid. Note that in our notation here p_ϕ is a projection of $\mathbf{p}$ on $\mathbf{e}_\phi$ and is related to the z-projection of the angular momentum as

$$(\mathbf{r}\times\mathbf{p})_z = rp_\phi = L_z$$

Compared to the situation without the magnetic flux the Hamiltonian (86) is modified by the presence of the potential A_ϕ in the centrifugal term which depends on the combination

$$p_\phi + eA_\phi = \frac{1}{r}(L_z + eBr_0^2/2) = \frac{1}{r}(L_z + e\Phi/2\pi)$$

In classical mechanics one could absorb the constant $e\Phi/2\pi$ into L_z and completely eliminate A_ϕ from the equations of motion. However in quantum mechanics this freedom does nor exist since L_z becomes an operator $L_z = -i\hbar\,\partial/\partial\phi$ which has discrete eigenvalues $\hbar M$ (M – integer). The eigenvalues' selection follows from the requirement that the wave function is single valued which imposes the periodic boundary conditions

$$\psi(r,\phi,z) = \psi(r,\phi+2\pi,z).$$

The eigenvalues of the angular part $L_z + e\Phi/2\pi$ in the expression for $p_\phi + eA_\phi$ are therefore

$$\hbar(M + e\Phi/2\pi\hbar) \equiv \hbar(M + \Phi/\Phi_0)\ \ ,\ \ \Phi_0 = h/e$$

which is identical with what was obtained in Eq. (82) of our discussion of the effect of the multivalued wave function condition obtained after "gauging out" the vector potential.

In the Schrödinger equation $H_{op}\psi = E\psi$ one can separate the z-part and use

$$\psi(r,\phi) = R(r)\frac{1}{\sqrt{2\pi}}e^{iM\phi}$$

to write the radial part of the equation as

$$\left[\frac{\hat{p}_r^2}{2m} + \frac{\hbar^2(M+\Phi/\Phi_0)^2}{2mr^2} + U(r)\right]R(r) = \varepsilon\,R(r)$$

with ε the corresponding part of the total energy E.

The above change of the spectrum of the centrifugal part of the potential is a manifestation of the Aharonov-Bohm effect in this example. A classically unobservable magnetic flux inside an impenetrable solenoid causes an observable effect in the outside region when the problem is treated quantum mechanically. The dependence on the magnitude Φ of the flux exhibits periodicity with magnetic flux quantum Φ_0 as a period.

7.4.2 Thin ring solution

Let us see how this happens in a simple model of a thin ring. To construct this model we add to the Hamiltonian (86) a potential $V(r, z)$ constraining the motion in the r and z directions to a very narrow ring region. It is the simplest to choose $V(r, z)$ as zero for $|r - a| \leq b$, $|z| \leq b$ and infinite otherwise. This gives a ring of thickness b with radius a lying in the $z = 0$ plane. For a very small b the radial coordinate in the second term in (86) can be set to the fixed radius a and the motion in the azimuthal direction becomes decoupled from r. The Hamiltonian of this motion is just

$$H_\phi = \frac{1}{2ma^2}\left(-i\hbar\frac{\partial}{\partial\phi} + \frac{e\Phi}{2\pi}\right)^2 \tag{87}$$

with eigenfunctions

$$\psi_M(\phi) = \frac{1}{\sqrt{2\pi}}\ \exp(iM\phi)\ \ ,\ \ M = 0, \pm1, \pm2, ...$$

and the corresponding eigenvalues

$$E_M = \frac{\hbar^2}{2ma^2}\left(M + \frac{\Phi}{\Phi_0}\right)^2\ . \tag{88}$$

The energies of the motion in the r and the z directions in this approximation are independent of Φ and we will not be concerned with them.

In Fig. 5 we plot the dependence of the energy levels on the magnetic flux which shows the Φ_0 periodicity of the Aharonov–Bohm effect. An analysis which we do not reproduce here shows that if there is a weak additional potential $V(\phi)$ acting on a particle on the ring the behavior of the levels will follow the pattern of the solid lines in Fig. 5 which retain the same periodicity, Ref.[13].

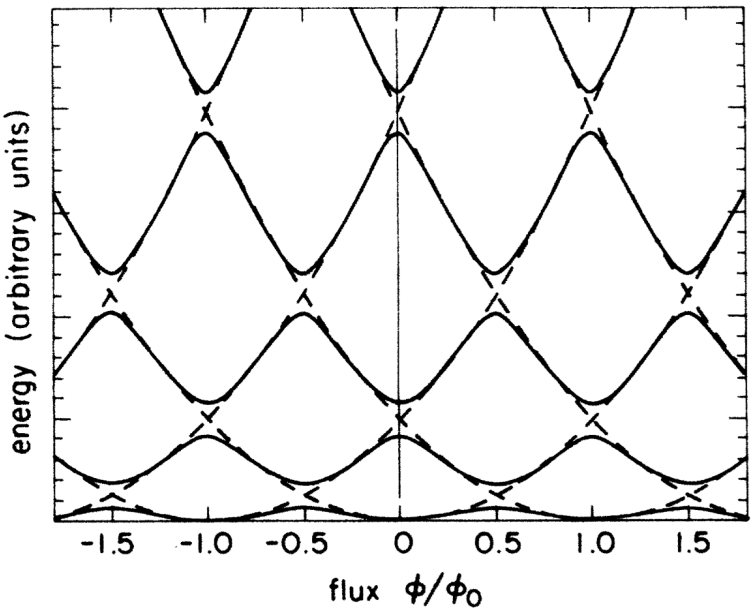

Figure 5: Schematic diagram of the electron energy levels as a function of the flux Φ/Φ_0 in a one-dimensional ring encircling the flux, Ref.[13]. Solid and dashed curves, respectively, are for the ring with and without weak additional potential $V(\phi)$ acting on a particle on the ring, Ref.[13].

Consider now a case of a strong potential $V(\phi)$, so strong that the particle is localized in a finite sector of the ring as opposed to the free motion around the entire circumference of the ring as in (87). A simple such $V(\phi)$ is a potential "well" $V(\phi) = 0$ for $0 < \phi < \phi_0 < 2\pi$ and infinite outside this interval. The eigenfunctions in this case are zero except in the interval with zero potential where they are easily found to be

$$\psi_n(\phi) = \sqrt{\frac{2}{\phi_0}} \exp\left[i\frac{\Phi}{\Phi_0}\,\phi\right] \sin\left(\frac{\pi n \phi}{\phi_0}\right) \quad , \quad n = 1, 2, 3, ... \tag{89}$$

The dependence on the flux enters in the phase of these functions but the corresponding eigenenergies do not depend on it at all,

$$E_n = \frac{\hbar^2 \pi^2 n^2}{2ma^2\phi_0^2}\,. \tag{90}$$

In Fig. 5 they would be represented by horizontal straight lines giving a trivial limiting case of the general periodic dependence on Φ referred to above. Here we have an

example in which the localization of the eigenfunctions on a part of the ring leads to the disappearance of the the Aharonov–Bohm effect (the Φ dependent phase is the same for all solutions and is therefore not observable in this case). As we have already stressed, in order to have a sensitivity to the Aharonov – Bohm flux the wave function must have a "tail" extending all around the flux.

7.4.3 AB effect in quantum interference and scattering off the AB flux

Here we will briefly consider two additional manifestations of the AB effect.

2-slit with AB flux

The understanding that Aharonov–Bohm flux modifies the phase of the wave function leads to an intuitive way of describing the Aharonov – Bohm effect as the change of the interference of the quantum mechanical waves as they propagate on each side of the solenoid. Let us consider the classic 2-slit experiment as depicted in Fig.6.

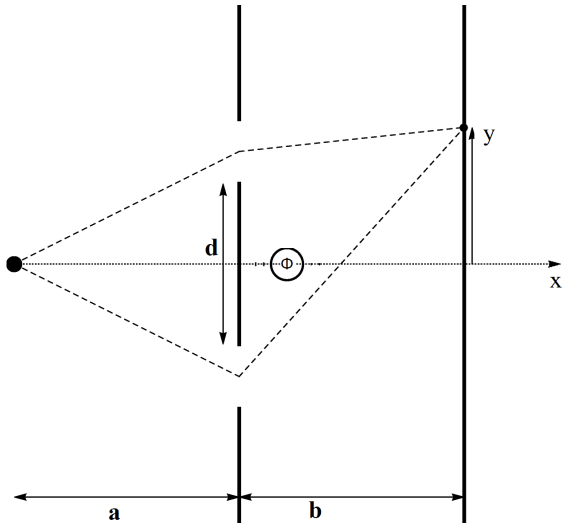

Figure 6: Double slit interference in the presence of the Aharonov-Bohm flux Φ

Electrons pass from a point source through a wall with 2 narrow slits and fall on a screen behind it, cf., Ref. [8]. As long as it is not detected which slit the electrons pass through, they produce an interference pattern according to the phase difference for paths going via each of the slits.

If the Aharonov–Bohm solenoid is placed behind the wall between the slits this

phase difference will change by the amount

$$\Delta\beta = \frac{e}{\hbar}\left(\int_1 \mathbf{A}\cdot d\mathbf{r} - \int_2 \mathbf{A}\cdot d\mathbf{r}\right) = \frac{e}{\hbar}\oint \mathbf{A}\cdot d\mathbf{r} = \frac{2\pi\Phi}{\Phi_0} \tag{91}$$

where as before Φ is the flux through the solenoid and the subscripts 1 and 2 denote integrals along the two trajectories in Fig. 6. For a position y on the screen (measured from its centre) the phase difference between waves from the two slits in the absence of the solenoid is $\beta = k\Delta L$ where $k = \sqrt{2mE}/\hbar$ is the wave number and ΔL – the difference in the paths the waves travel from the slits to the screen.

For a distance b from the slits to the screen and for $y, d << b$ one can approximate $\Delta L = (y/b)d$ where d is the distance between the slits. Therefore a given phase difference β will be found at $y = (\beta/kd)b$. The additional phase difference $\Delta\beta$ due to the Aharonov – Bohm flux will result in a shift in the interference pattern by the amount

$$\Delta y = \frac{\Delta\beta}{kd}b = \frac{2\pi b}{kd}\frac{\Phi}{\Phi_0} \tag{92}$$

Scattering off the AB flux

Yet another way to see the phase difference between the waves which pass on different sides of the solenoid is to consider a scattering of a plane wave from it. This was discussed in the original paper by Aharonov and Bohm, Ref. [7]. For the vanishing magnetic flux one finds a standard picture of a cylindrical wave scattered from the solenoid with the amplitude which falls like r^{-1} superimposed on the initial plane wave. For non zero Φ the wavefronts in the region "down stream" and far away from the solenoid form a pattern of two flat fronts shifted with respect to each other in an abrupt, almost discontinuous fashion along the line stretching from the solenoid in the direction of the propagation of the original plane wave. The magnitude of the shift is given by the phase difference $2\pi\Phi/\Phi_0$ divided by the wave number k. We refer the reader to Ref. [7] for the details of this discussion.

7.5 Multiply connected regions. Homotopy

Certain features of the results obtained in examples above are quite general in nature. In any region with zero $\mathbf{E}$ and $\mathbf{B}$ Eqs. (2) imply that the vector potential must be the gradient of a (time-independent) function, $\mathbf{A}(\mathbf{r}) = \nabla\xi(\mathbf{r})$ and that $A_0 = const.$ The integral $\int \mathbf{A}\cdot d\mathbf{r} = \int \nabla\xi \cdot d\mathbf{r}$ is equal to the difference between the values of ξ at the initial and the final points of the contour of the integration so that it must be zero for a closed contour unless (a) the function ξ is *not single valued* and (b) the contour takes ξ from one of its branches to another. This can not happen in simply connected regions, i.e. such in which *all closed contours are contractable to a point.* Since continuous deformations of the contour in the integral $\oint \nabla\xi \cdot d\mathbf{r}$ can not change its value all such integrals will be zero for contractable contours. Equivalently stated, a regular function like $\xi(\mathbf{r})$ must be single valued in a simply connected region.

Consider however multiply connected regions. These are regions where one can find closed contours which can not be contracted to a point without crossing the boundaries. The impenetrable solenoid and the ring discussed above are examples of such regions. The contours around the "tube" of the solenoid or around the "hole" of the ring can not be contracted to zero. Non zero values for the circulation integral $\int \mathbf{A}\cdot d\mathbf{r}$ are to be expected for such contours and actually occur when there is a magnetic flux through the excluded regions. We note in passing that not every shape of excluded region will lead to the existence of non contractable contours. Excluded cavity of a spherical shape for instance will not. Its existence will create uncontractable closed *surfaces but not curves* and will be relevant for considerations of e.g. non vanishing surface integrals of a vector field with zero divergence.

All possible closed curves in a multiply connected region can be divided into classes

of curves which can be contracted into each other. Such classes are called homotopy classes of curves, cf., [9]. Among all homotopy classes one can define a complete set of elementary classes of curves C_k out of which every other non elementary class can be obtained by multiple traverses of curves belonging to the elementary classes. For a solenoid there is one elementary class of curves encircling the solenoid once in, say, a clockwise direction and one in the counter clockwise direction. Clearly the changes of ξ on closed curves within each elementary class must be the same since the curves can be continuously deformed into each other. For different elementary classes however they in general will be different reflecting possible different values of the Aharonov–Bohm fluxes through different excluded regions or their different signs.

Let us apply these considerations to a general system of charged particles in a multiply connected field free region, cf. Ref. [10]. Their Schrödinger equation is

$$H[\mathbf{p}_a - q_a\mathbf{A}(\mathbf{r}_a),\ \mathbf{r}_a]\psi(\{\mathbf{r}_a\}) = E\psi(\{\mathbf{r}_a\}) \tag{93}$$

where we assumed a general Hamiltonian depending on the momenta $\mathbf{p}_a = -i\hbar\nabla_a$ and coordinates $\mathbf{r}_a$ of the particles with charges $q_a,\ a = 1, 2, ..., N$. Since in the field free region the vector potential $\mathbf{A}$ is "pure gauge", $\mathbf{A} = \nabla\xi$, we can apply a gauge transformation

$$\psi(\{\mathbf{r}_a\}) = \exp[i\sum_{a=1}^{N} q_a\xi(\mathbf{r}_a)/\hbar]\psi'(\{\mathbf{r}_a\}) \tag{94}$$

and find that ψ' satisfies

$$H[\mathbf{p}_a, \mathbf{r}_a]\psi'(\{\mathbf{r}_a\}) = E\psi'(\{\mathbf{r}_a\}) \tag{95}$$

with the Hamiltonian in which the potential $\mathbf{A}$ was "gauged out". Since ψ is single valued and since going around any elementary non contractable contour C_k increases ξ by the corresponding Aharonov–Bohm flux $\Phi_k = \oint_{C_k} \mathbf{A}\cdot d\mathbf{r}$, we must demand that $\psi'(\{\mathbf{r}_a\})$ is multiplied by the factor $\exp(-iq_a\Phi_k/\hbar)$ when the particle a is brought around C_k. Thus the boundary conditions are different from the case of zero fluxes and one should expect that the energy levels will depend on the values of Φ_k. Since charges of all particles are multiples of the elementary electronic charge e the change in the boundary conditions is the same for the fluxes Φ_k which differ by multiples of the flux quantum Φ_0. This periodicity should occur in the solutions ψ' and therefore in the set of energy levels obtained from Eq.(94). Physical quantities which are determined by this set must therefore exhibit this periodicity. This conclusion as well as the entire set of the preceding arguments are very general and based solely on the fundamental principles of gauge invariance, requirement of single valued wave functions and the elementary nature of the electric charge e.

8 Magnetic Moments

8.1 The g–factors

We now return to the relation (10) between the magnetic moment and the spin operators. It is customary to quote the numerical value of the magnetic moment of a particle as equal to the maximum value of its projection, i.e. the value $\mu_z = g(q/2mc)s_z$ for $s_z = s$. In this Section we will discuss the dimensionless gyromagnetic ratio g in this relation called the g-factor.

For *elementary particles* g is determined by relativistic quantum mechanical wave equations. E.g. for the electron the Dirac equation gives $g = 2$, i.e. *twice the classical value*. Unlike orbital angular momentum or the spin of a composite particle the spin of an elementary particle has a fixed value and therefore its magnetic moment is *fixed* and must be regarded as one of the characteristics of the particle like its charge, mass, etc. The electron magnetic moment (spin 1/2) is to a good approximation given by the Dirac value[7]

$$\mu_0 = \frac{\mid e \mid \hbar}{2mc} \, . \tag{96}$$

This quantity is called the *Bohr magneton* and serves as a convenient unit in which magnetic moments are measured in atomic physics. Its numerical value is $5.79 \cdot 10^{-9}$ eV/Gauss.

In nuclear physics a more appropriate unit is the *nuclear magneton* defined as in (96) but with the mass of the proton used for m. Experimentally measured value for the magnetic moment of the proton is 2.793 nuclear magnetons meaning that the g-factor is 5.586. For neutrons the values are –1.913 and –3.826 respectively. The deviation of these g-factors from the corresponding Dirac values $g = 2$ and $g = 0$ was among the first experimental indications that protons and neutrons are not elementary but rather composite particles. In general the calculation of the g-factors for composite particles requires the knowledge of the intrinsic dynamics, i.e. the wave function of the elementary constituents, their spins, etc. We will consider examples of such calculations below.

[7]Small deviations from this value are very accurately described in Quantum Electrodynamics by the effects of the interaction with the surrounding cloud of virtual photons and electron–positron pairs.

8.2 Atoms in a magnetic field

8.2.1 The Hamiltonian

Consider an atom placed in a uniform magnetic field. Assuming fixed heavy nucleus atomic electrons are described by the Hamiltonian[8]

$$H = \frac{1}{2m}\sum_a \left(\mathbf{p}_a + \frac{e}{c}\mathbf{A}(\mathbf{r}_a)\right)^2 + U(\mathbf{r}_a) + \frac{e}{mc}\mathbf{B}\cdot\sum_a \mathbf{s}_a \ , \tag{97}$$

where we denoted by $\mathbf{r}_a$, $\mathbf{p}_a$ and $\mathbf{s}_a$ the coordinates, momenta and spin operators of the electrons and included in $U(\mathbf{r}_a)$ the interaction of the electrons with the atomic nucleus as well as their Coulomb interaction with each other. We used $q = -e$ for electrons and for simplicity disregarded the nuclear spin.

Choosing the vector potential in the form $\mathbf{A} = (\mathbf{B}\times\mathbf{r})/2$ for which $\nabla\cdot\mathbf{A} = 0$ we can write the Hamiltonian in the form

$$\begin{aligned} H &= H_0 + \frac{e}{2mc}\mathbf{B}\cdot\sum_a(\mathbf{r}_a\times\mathbf{p}_a) + \frac{e^2}{8mc^2}\sum_a(\mathbf{B}\times\mathbf{r}_a)^2 + \frac{e}{mc}\mathbf{B}\cdot\sum_a\mathbf{s}_a \qquad (98)\\ &= H_0 + \mu_0(\mathbf{L}+2\mathbf{S})\cdot\mathbf{B} + \frac{e^2}{8mc^2}\sum_a(\mathbf{B}\times\mathbf{r}_a)^2 \ , \qquad (99)\end{aligned}$$

where H_0 is the Hamiltonian in the absence of the magnetic field, μ_0 is the Bohr magneton and we used the expressions $\mathbf{L} = \sum_a(\mathbf{r}_a\times\mathbf{p}_a)$ and $\mathbf{S} = \sum_a \mathbf{s}_a$ for the total orbital angular momentum and spin. The terms in H which depend on the magnetic field can be written as $-\boldsymbol{\mu}\cdot\mathbf{B}$ with the operator of the magnetic moment

$$\boldsymbol{\mu} = -\mu_0(\mathbf{L}+2\mathbf{S}) + \frac{e^2}{8mc^2}\sum_a[\mathbf{r}_a^2\mathbf{B} - \mathbf{r}_a(\mathbf{r}_a\cdot\mathbf{B})] \ . \tag{100}$$

The first term in this expression is independent of $\mathbf{B}$ and can be considered as the operator of the intrinsic magnetic moment of the atom which exists in the absence of the field. It is a sum of the orbital and the spin contributions in which the latter enters with twice as large coefficient. It is crucial to observe that because of this non classical Dirac value of the spin g–factor the intrinsic magnetic moment is not parallel to the system total angular momentum $\mathbf{J} = \mathbf{L} + \mathbf{S}$. As we will presently see this is the main reason why in general the atomic g–factors do not have the universal classical value $g = 1$ but depend on the state of the atom.

[8]In this and the following Sections we use CGS units

The second term in $\boldsymbol{\mu}$ depends on $\mathbf{B}$ and must be regarded as the operator of the magnetic moment which is *induced* by the magnetic field. Its magnitude $-(e^2/8m^2c^2)\sum_j I_{ij}B_j$ is proportional to the moment of inertia $I_{ij} = \sum_a m(r_{a,i}r_{a,j} - \delta_{ij}\mathbf{r}_a^2)$ which of is one of the manifestations of the Larmor theorem.

8.2.2 Treating the B dependent terms perturbatively. LS and jj couplings

Exact diagonalization of the Hamiltonian (99) is not feasible even when the solutions in the absence of the magnetic field are known. The standard way of treating this problem is to use the perturbation theory with respect to the B–dependent terms. Let us start with the linear term in (99). Because of the rotational symmetry the states of the atom are characterized by the eigenvalues $J(J+1)$ of $\mathbf{J}^2$ and for non zero J are multiplets of degenerate states which can be labeled by one of the projections of $\mathbf{J}$. One must therefore use degenerate perturbation theory and to lowest order diagonalize the perturbation $H_1 = \mu_0(\mathbf{L} + 2\mathbf{S}) \cdot \mathbf{B}$ in the subspace of each multiplet. The magnetic field breaks the rotational symmetry and removes the multiplet degeneracies. The remaining axial symmetry of rotations around the direction of $\mathbf{B}$ indicates that within each multiplet of the degenerate states the correct combinations which diagonalize H_1 are the eigenstates of the projection of $\mathbf{J}$ on $\mathbf{B}$. The energy shift of these states with respect to the unperturbed value is simply the expectation value of H_1,

$$\Delta E = \mu_0 B < \alpha; J, M|L_z + 2S_z|\alpha; J, M >= \mu_0 B < \alpha; J, M|J_z + S_z|\alpha; J, M > , \quad (101)$$

where we have chosen the z-axis along the direction of $\mathbf{B}$ and denoted by α the additional quantum numbers apart of J and its projection M which are needed in order to specify an atomic state.

According to the Wigner–Eckart theorem, cf. Ref. [12], the matrix element of a component of any vector operator between states of a multiplet with a given J is proportional to the same matrix element of the same component of the operator $\mathbf{J}$ with the proportionality constant which is independent of M. We can therefore write

$$\Delta E = \mu_0 g_{\alpha,J} B < \alpha; J, M|J_z|\alpha; J, M >= \mu_0 g_{\alpha,J} BM , \quad (102)$$

where the yet undetermined proportionality constant $g_{\alpha,J}$ obviously represents the g–factor of the atomic state. Finding explicit expression for $g_{\alpha,J}$ requires further information about the structure of $|\alpha, J, M >$ and can only be made in certain limiting cases.

If the interactions in atoms were the ordinary Coulomb forces the total orbital and spin angular momenta and their projections M_L and M_S would be separately conserved and in this case

$$\Delta E = \mu_0 B(M_L + M_S) \; . \tag{103}$$

In reality, however relativistic effects are important and produce the so called *fine structure* of atomic levels. The main relativistic effect turns out to be the presence in the atomic Hamiltonian H_0 of the spin–orbit term $\sum_a V_{so}(|\mathbf{r}_a|)\; \mathbf{l}_a \cdot \mathbf{s}_a$ with $V_{so}(r)$ proportional to r^{-1} times the derivative with respect to r of the atomic potential $-Ze^2/r$. When this term is relatively weak (as happens for most atomic states) one can treat it as a perturbation and diagonalize it separately within each degenerate multiplet of $(2L+1)(2S+1)$ states with given L and S.

The result is what is called the *fine splitting* of the multiplet into closely lying states which have definite values of J. In this zero order of the perturbation treatment they are linear combinations of the unperturbed wave functions with same values of L and S but different M_L and M_S. Formally these *zeroth order* atomic states are written as

$$|n; L, S; J, M >= \sum_{M_L + M_S = M} < L, M_L; S, M_S | L, S; J, M > |n, L, M_L > |S, M_S >$$

and are referred to as states of the "LS – coupling" scheme. By n we denoted here the remaining quantum numbers for the orbital motion and the coefficients in the sum are the standard Clebsh-Gordan coefficients for coupling of two angular momenta.

In the opposite extreme case of the strong spin–orbit interaction one can not talk about separate conservation of the orbital and spin angular momenta. Individual electrons must be characterized by their total angular momenta j which must be combined to produce the total J. Such a scheme of constructing the zeroth order wave functions is called the "jj – coupling". This extreme limit is rarely found in atoms but plays a central role in nuclear spectroscopy.

8.2.3 Lande formula

For states with LS – coupling a general expression for the g-factors called the Lande formula can be derived,

$$g = 1 + \frac{J(J+1) - L(L+1) + S(S+1)}{2J(J+1)} \; . \tag{104}$$

This is found as follows. As was already mentioned the Wigner – Eckart theorem gives

$$< \mathbf{S} >= const \cdot < \mathbf{J} > \tag{105}$$

where we use the angular brackets to denote averages with respect to the state $|n; L, S; J, M >$. Since the operator $\mathbf{J}$ commutes with $\mathbf{L}$ and $\mathbf{S}$ it does not change the quantum numbers of this state so we can write

$$< \mathbf{S} \cdot \mathbf{J} >= const < \mathbf{J} \cdot \mathbf{J} >$$

with the same constant. Using $< \mathbf{J} \cdot \mathbf{J} >= J(J+1)$ have

$$< S_z >= const\ M = M \frac{< \mathbf{S} \cdot \mathbf{J} >}{J(J+1)} \tag{106}$$

Using $\mathbf{L} \cdot \mathbf{L} = (\mathbf{J} - \mathbf{S})^2 = \mathbf{J} \cdot \mathbf{J} + \mathbf{S} \cdot \mathbf{S} - 2\mathbf{J} \cdot \mathbf{S}$ and the properties of the LS – coupling state we find that

$$< \mathbf{S} \cdot \mathbf{J} >= \frac{1}{2}[J(J+1) - L(L+1) + S(S+1)] \tag{107}$$

Collecting the results in Eq. (101) we obtain that ΔE is in the form (102) with $g_{\alpha,J}$ given by the Lande expression (104). As usual with the results of the perturbation theory this formula is valid when ΔE are small as compared to the intervals between the unperturbed atomic energy levels. In the present case these are the intervals due to the fine structure splitting.

8.3 The Zeeman effect

The general phenomenon of the energy splitting of atomic levels in magnetic field is called the Zeeman effect. The Lande formula gives the classical value $g = 1$ in the case $S = 0$ and the Dirac value $g = 2$ when $L = 0$. Historically the measured deviations of g from the classical value 1 were termed the anomalous Zeeman effect. In the case when the magnetic field is so intense that $\mu_0 B$ is larger than the intervals of the fine structure the energy splittings ΔE deviate from the predictions of the Lande formula. This is called the Pashen – Back effect. We will not discuss the details of it.

Let us now turn to the last term in the Hamiltonian (99) which is quadratic and describes as we already mentioned the interaction of induced magnetic moment with the field $\mathbf{B}$. This interaction is sometimes called diamagnetic to distinguish it from

the linear term which is called the paramagnetic interaction . The relative magnitude of these two terms can be estimated as $(e/\hbar c)\, r^2 B \sim 4 \cdot 10^6 (r/\mathrm{cm})^2 B/Gauss$ and one finds that for typical magnetic fields in laboratory the diamagnetic term is negligible if r has atomic dimensions. However when an atomic state has zero spin and orbital angular momentum ($L = S = 0$), the linear term does not effect the energy levels in any order of the perturbation since it has vanishing matrix elements. In this case the entire effect is determined by the quadratic term. In first order of the perturbation theory the corresponding energy shift is

$$\Delta E = \frac{e^2}{8mc^2} \sum_a < (\mathbf{r}_a \times \mathbf{B})^2 > , \tag{108}$$

where the average is with respect to a (non degenerate) state with L = S = 0. Since $< (\mathbf{r}_a \times \mathbf{B})^2 >= B^2 < r_a^2 \sin^2\theta_a >$ and since the wave function of a state with L = S = 0 is spherically symmetric one can average first over the angle and obtain $< (\mathbf{r}_a \times \mathbf{B})^2 >= 2B^2 < r_a^2 > /3$ where we used

$$< \sin^2\theta_a >= \int \sin^2\theta_a \; 2\pi \cos\theta_a \; d\theta_a / 4\pi = 2/3$$

Therefore

$$\Delta E = \frac{e^2}{12mc^2} B^2 \sum_a < r_a^2 > \tag{109}$$

Having in mind the general expression $-\boldsymbol{\mu} \cdot \mathbf{B}$ we see that the change of the induced magnetic moment with the field in this case is negative which means that such a state is diamagnetic.

9 Time Reversal in Magnetic Field. Kramers Degeneracy

In the absence of magnetic field and for spinless particles the Schrödinger equation with a time independent Hamiltonian is invariant under the substitution $t \to -t$ provided one also changes $\psi \to \psi^*$. One adopts

$$\psi(\mathbf{r},\mathrm{t}) \to T\psi(\mathbf{r},\mathrm{t}) \equiv \psi^*(\mathbf{r}, -t) \tag{110}$$

as a definition of the time reversal transformation in this case. Magnetic field and the particle spin require modifications of this definition. Since magnetic field acts also on the spin variables it is natural to discuss them together.

Even time independent magnetic field breaks the time reversal symmetry. This is already known in classical physics. The equation of motion (4) is time reversal invariant for any static $\mathbf{E}$ if $\mathbf{B} = 0$. For non vanishing $\mathbf{B}(\mathbf{r})$ this symmetry is lost but one observes that the equation retains its form if together with the sign change of t one changes the sign of the magnetic field. Thus all solutions $\mathbf{r}(t)$ found in a given $\mathbf{E}(\mathbf{r})$ and $\mathbf{B}(\mathbf{r})$ must have "partners" in the form $\mathbf{r}(-t)$ in a related problem with $\mathbf{E}(\mathbf{r})$ and $-\mathbf{B}(\mathbf{r})$. Of course one must take care in defining properly matched initial conditions for related solutions, i.e. impose time reversed initial velocities. One easily understands why the sign of $\mathbf{B}$ must be reversed – this is consistent with Maxwell equations which relate $\mathbf{B}$ to external currents which change their direction under time reversal. Similar arguments make it clear why $\mathbf{E}$ should stay the same.

Let us now turn to quantum mechanics in a static electromagnetic field. We first notice that changing the sign of t and of $\mathbf{B}$ without changing $\mathbf{E}$ simply means that $\mathbf{A} \to -\mathbf{A}$ together with $t \to -t$. Transforming also $\psi(\mathbf{r}, t) \to \psi^*(\mathbf{r}, -t)$ in the Schrödinger equation (12) we see that such a combined transformation leaves invariant all the terms in the equation except for the last, spin dependent term which becomes $ge\mathbf{B} \cdot \mathbf{s}^*\psi^*/2mc$ rather than $-ge\mathbf{B} \cdot \mathbf{s}\psi^*/2mc$. By analogy with the orbital angular momentum one needs the reversal of the sign of the spin operators and the complex conjugation does not accomplish this. Indeed recalling the standard representation of the spin operators in terms of the Pauli matrices,

$$s_x = \frac{\hbar}{2}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \qquad s_y = \frac{\hbar}{2}\begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \qquad s_z = \frac{\hbar}{2}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

one sees that (*in this particular representation*, in which s_z is diagonal) the complex conjugation causes only $s_x^* = s_x$, $s_y^* = -s_y$ and $s_z = s_z^*$. Hence one must together with the complex conjugation also change the sign of s_x and s_z without changing s_y. This can be accomplished by the rotation by the angle π around the y–axis in the "space" of the spin variables. Simce $\mathbf{s}$ is the operator of infinitesimal rotations in this space such a rotation is achieved by the operator $\exp[\, i\pi s_y/\hbar]$. Accordingly, we generalize the time reversal transformation of the wave function for particles with spin as

$$\psi(\mathbf{r},\mathrm{t}) \to T\psi(\mathbf{r},\mathrm{t}) \equiv \exp[\, i\pi s_y/\hbar]\psi^*(\mathbf{r}, -t)\ , \tag{111}$$

which must be supplemented with the sign change of $\mathbf{A}$ and $\mathbf{B}$ in the presence of the magnetic field. Now of course all the terms in Eq.(12) will transform correctly. We note that the transformation (111) (as well as the incomplete (110)) is antilinear, i.e. $T(\alpha\psi_1 + \beta\psi_2) = \alpha^* T\psi_1 + \beta^* T\psi_2$ and antiunitary, i.e. $< T\psi|T\phi >=< \psi|\phi >^*$.

The transformation properties of any (possibly time dependent) quantum mechanical operator O_{op} under time reversal are determined by considering

$$T(O_{op}\psi) = \exp\left[\, i\pi s_y/\hbar\right] O^*_{op}(-t)\psi^*(\mathbf{r}, -t)$$

and comparing with $T(O_{op}\psi) = (T\, O_{op}\, T^{-1})(T\psi)$. We thus find

$$T\,\mathbf{r}\,T^{-1} = \mathbf{r}\ \ ,\ \ T\,\mathbf{p}\,T^{-1} = -\mathbf{p}\ \ ,\ \ T\,\mathbf{s}\,T^{-1} = -\mathbf{s}\,, \tag{112}$$

in line with the intuition.

It is important to remember that the explicit form of the time reversal operator as given above was derived in the particular representation, i.e. in the coordinate representation and diagonal spin projection s_z. It is in general not valid in other representations, but can be derived following the rules of transformations between representations. E. g., a plane wave $\exp(i\mathbf{k}\cdot\mathbf{r})$ in the coordinate representation becomes $\delta(\mathbf{p}-\hbar\mathbf{k})$ in the momentum representation for which the complex conjugation is obviously not producing what expected under the time reversal – the change of the sign of $\mathbf{k}$. Using the relation between the coordinate and momentum representations we obtain

$$< \mathbf{p} \mid T\psi >= \int d\mathbf{r} < \mathbf{p} \mid \mathbf{r} >< \mathbf{r} \mid \psi(-t) >^* =< -\mathbf{p} \mid \psi(-t) >^*\,, \tag{113}$$

where we ignored the spin and used $< \mathbf{p} \mid \mathbf{r} >=< -\mathbf{p} \mid \mathbf{r} >^*$. It is seen that the time reversal in momentum representation is a combined action of the complex conjugation and the change of sign of the momenta — not surprising.

Let us return to physical systems without external magnetic field. Their Hamiltonians are symmetric under time reversal, $[H, T] = 0$. For an eigenstate ψ_n of H this gives $HT\psi_n = TH\psi_n = E_n\psi_n$ which means that ψ_n and $T\psi_n$ have the same energy. There are two strong results which follow from this fact :

- for spinless particles non degenerate eigenstates of H can always be chosen to be real and
- eigenstates with *half–integer total spin* are always at least doubly degenerate. This degeneracy is called *the Kramers degeneracy.*

To prove the first result we note that since for spinless particles $T\psi_n(\mathbf{r}) = \psi^*_n(\mathbf{r})$ and since by the assumption E_n is not degenerate the function $\psi_n(\mathbf{r})$ must coincide with $\psi^*_n(\mathbf{r})$ up to a constant independent of $\mathbf{r}$. For normalized wave functions this

is at most a phase factor which is inessential for any physical results and can be disregarded.

In order to prove the second result consider an eigenstate wave function $\psi_n = \psi_{\alpha jm}$ and its time reversed partner $T\psi_{\alpha jm}$, where we denoted by jm the total spin of the system and its projection and by α all other quantum numbers. Should these functions represent the same state as in the spinless case they would be related as $T\psi_{\alpha jm} = C\psi_{\alpha jm}$ with some complex constant C. Applying T once again we would get $T^2\psi_{\alpha jm} = |C|^2\psi_{\alpha jm}$. But on the other hand $T^2 = \exp[\,2i\pi s_y/\hbar]$ which gives $(-1)^{2j}$ when applied to $\psi_{\alpha jm}$, cf., Problem 1. This can not be equal to a positive $|C|^2$ for half–integer j. We are therefore led to conclude that $\psi_{\alpha jm}$ and $T\psi_{\alpha jm}$ must correspond to different states for half–integer j which means that the corresponding eigenvalue E_n is at least doubly degenerate. This Kramers degeneracy means, for instance, that for a system with odd number of electrons the energy levels will always be at least twofold degenerate even if it is placed in any, however complicated, electric field

10 Path Integrals with the External Electromagnetic Field

The derivation of the path integral quantization of a particle in the presence of the electromagnetic field follows the standard route. The propagator

$$K(\mathbf{r}_f, t_f; \mathbf{r}_i, t_i) \equiv \langle \mathbf{r}_f | e^{-iH_{op}(t_f - t_i)} | r_i \rangle$$

is represented as a multiple integral

$$\begin{aligned} K(\mathbf{r}_f, t_f; \mathbf{r}_i, t_i) &= \lim_{N\to\infty} \int d^3 r_N \int d^3 r_{N-1} \ldots \int d^3 r_1 \; K(\mathbf{r}_f, t_f; \mathbf{r}_N, t_N) \times \\ &\quad \times \; . \; . \; . \times K(\mathbf{r}_1, t_1; \mathbf{r}_i, t_i) \; , \end{aligned} \tag{114}$$

over infinitesimal propagators which should be calculated for the Hamiltonian operator given by Eq. (8) (we do not consider the spin dependent term – such terms require special treatment in the path integral formulation). Based on the experience with path integrals one should expect that the propagator $K(\mathbf{r}, t+\epsilon; \mathbf{r}', t)$ over an infinitesimal time interval ϵ is expressed as

$$\Big(\frac{m}{2\pi i\hbar\epsilon}\Big)^{3/2} \exp\{\frac{i}{\hbar}\epsilon L[(\mathbf{r}+\mathbf{r}')/2, (\mathbf{r}-\mathbf{r}')/\epsilon]\}$$

in terms of the classical Lagrangian $L(\mathbf{r}, \mathbf{v})$ given by Eq. (5). An explicit calculation indeed shows that

$$\psi(\mathbf{r}, t+\epsilon) = \int d\mathbf{r}' K(\mathbf{r}, t+\epsilon; \mathbf{r}', t)\, \psi(\mathbf{r}', t) \tag{115}$$

reproduces the Schrödinger equation with the infinitesimal propagator given by

$$\begin{aligned} K(\mathbf{r}, t+\epsilon; \mathbf{r}', t) &= \left(\frac{m}{2\pi i\hbar\epsilon}\right)^{3/2} \exp\left\{\frac{i\epsilon}{\hbar}\left[\frac{m}{2}\left(\frac{\mathbf{r}-\mathbf{r}'}{\epsilon}\right)^2 - eA_0\left(\frac{\mathbf{r}+\mathbf{r}'}{2}\right)\right] + \right. \\ &\left. + \frac{ie}{\hbar c}(\mathbf{r}-\mathbf{r}')\cdot\mathbf{A}\left(\frac{\mathbf{r}+\mathbf{r}'}{2}\right)\right\} \end{aligned} \tag{116}$$

The details of this calculation are rather cumbersome and will not be reproduced here. They can be found in Ref. [4].

Using the expression for the infinitesimal propagator in the multiple integral for $K(\mathbf{r}_f, t_f; \mathbf{r}_i, t_i)$ we find after combining the product of the exponentials into a exponential of a sum and using the continuous notation

$$K(\mathbf{r}_f, t_f; \mathbf{r}_i, t_i) = \int_{\mathbf{r}(t_i)=\mathbf{r}_i}^{\mathbf{r}(t_f)=\mathbf{r}_f} D[\mathbf{r}(t)] \exp\left\{\frac{i}{\hbar}\int_{t_1}^{t_2} dt\left[\frac{m\mathbf{v}^2}{2} - eA_0(\mathbf{r}) + \frac{e}{c}\mathbf{A}(\mathbf{r})\cdot\mathbf{v}\right]\right\} \tag{117}$$

where as usual the definition of $D[\mathbf{r}(t)]$ includes the product of N d^3r_i's each multiplied by the pre-exponential factors from Eq. (116).

The first two terms in Eq. (117) are the usual kinetic and potential energies but the last term is a new feature of this path integral. It mixes coordinates and velocities but its linear dependence on $\mathbf{v}$ is special. Making the replacement $\mathbf{v}\cdot dt = d\mathbf{r}$ the contribution of this term for every path in the path integration can be written as

$$\exp\left[\frac{ie}{\hbar c}\int_{\mathbf{r}_i}^{\mathbf{r}_f} \mathbf{A}[\mathbf{r}(t)]\cdot d\mathbf{r}\right] \ .$$

This phase factor depends on the path but not on the velocity of propagation along it. If one considers a difference of these phases between two arbitrary paths one can write it as the circulation of $\mathbf{A}$ along a closed path which is obtained by traversing from $\mathbf{r}_i$ to $\mathbf{r}_f$ along one path and then back to $\mathbf{r}_i$ along the other. Using the Stokes theorem $\oint \mathbf{A}\cdot d\mathbf{r} = \int \nabla\times\mathbf{A}\cdot d\mathbf{S} = \int \mathbf{B}\cdot d\mathbf{S}$ one can write this phase difference as

$$\exp\left[2\pi\, i\, \frac{\Phi}{\Phi_0}\right] \ , \tag{118}$$

where Φ is the flux of the magnetic field through the closed contour defined by the two paths and $\Phi_0 = hc/e$ is the magnetic flux quantum already familiar from our discussions of the Aharonov-Bohm effect.

We would like to point out an important subtlety related to the appearance of terms like $\mathbf{A}(\mathbf{r}) \cdot \mathbf{v}$ in the path integration. One will get different answers depending on whether $\mathbf{A}(\mathbf{r})$ is evaluated at $(\mathbf{r}+\mathbf{r}')/2$, at $\mathbf{r}$, at $\mathbf{r}'$ or somewhere in between in the infinitesimal propagator (116). This ambiguity is known as the Ito ambiguity and is discussed in detail in Ref. [4]. It is shown there that the correct prescription is to take $\mathbf{A}$ as it is written in Eq.(116), i.e. *at a midpoint*. This is sometimes referred to as the *mid–point rule*. Only with this rule the correct Schrödinger equation is reproduced. The mid–point rule is important for a term $A\cdot\mathbf{v}$ and not for the conventional potential term $eA_0(\mathbf{r})$. This is because of *the different powers of* ϵ, i.e. ϵ^0 and ϵ^1 which multiply the discretized version of $\mathbf{A}\cdot\mathbf{v}$ and eA_0 respectively in the expression for the infinitesimal propagator. As is shown in standard discussions of the path integrals the typical distances between propagation points obey the estimate $|\mathbf{r}-\mathbf{r}'| \sim \sqrt{\epsilon}$. Changes of this order of magnitude in the argument of $eA_0[(\mathbf{r}+\mathbf{r}')/2]$ combined with ϵ^1 in front of it will contribute a negligible difference of the order $\sim \epsilon^{3/2}$. The same change in $\mathbf{A}[(\mathbf{r}+\mathbf{r}')/2]$ combined with the term $\mathbf{r}-\mathbf{r}' \sim \epsilon^{1/2}$ which multiplies it contributes $O(\epsilon^1)$ which can not be neglected.

Let us now examine how the gauge transformations effect the path integral (117). Performing a gauge transformation (3) of the potentials $\mathbf{A}$ and A_0 adds in the action the term proportional to

$$\int_{t_i}^{t_f} dt \left[\frac{d\mathbf{r}}{dt}\cdot\nabla\chi(\mathbf{r},\mathrm{t}) + \frac{\partial\chi(\mathbf{r},\mathrm{t})}{\partial t}\right] = \int_{t_i}^{t_f} \frac{d\chi(\mathbf{r},\mathrm{t})}{dt}dt = \chi(\mathbf{r}_f,t_f) - \chi(\mathbf{r}_i,t_i)\ , \qquad (119)$$

where the last equality holds because of the mid–point rule of the discretization of the integral and gives the result which is the same for all paths. Using this we find that under the gauge transformation the propagator changes as

$$K'(\mathbf{r}_f,t_f;\mathbf{r}_i,t_i) = \exp[ie\chi(\mathbf{r}_f,t_f)/\hbar c]K(\mathbf{r}_f,t_f;\mathbf{r}_i,t_i)\exp[-ie\chi(\mathbf{r}_i,t_i)/\hbar c]\ . \qquad (120)$$

This of course is of the same origin as the change of the phase of the wave function (16). The phase change of $K(\mathbf{r}_f,t_f;\mathbf{r}_i,t_i)$ depends only on the initial and the final coordinates. The phase difference between different paths is strictly gauge invariant.

In a *uniform* electric and magnetic field the Lagrangian (5) is a quadratic function of the coordinates and velocities and the path integral in this case is of the Gaussian type and can be evaluated exactly (cf., Problem 2 at the end of the Chapter).

11 Dirac Magnetic Monopoles

11.1 Multivalued wave functions. Non integrable phases

An instructive discussion related in a surprising way to the general issue of the gauge transformations arises when one examines in depth the requirement that the solutions of the Schrödinger equation must be single valued. This requirement is usually imposed as natural and is the main reason for finding the standard quantized values of physical quantities such as the energy, the angular momentum, etc. Following the discussion by Dirac [9] let us try to see what happens if this requirement is removed.

Of course one still must obtain unambiguous results for quantities which have direct physical meaning. This certainly means that the amplitude of the wave function must be single valued since its square is a physical density function. The phase of the wave function on the other hand does not have to have a unique value at a particular point so in general the wave function can be written as $\psi(\mathbf{r},\mathrm{t}) = \phi(\mathbf{r},\mathrm{t})\exp[i\beta]$ with $\phi(\mathbf{r},\mathrm{t})$ the ordinary single valued complex function and all multivaluedness residing in the properties of the phase β. A useful way to characterize this multivaluedness is to consider how β changes when one goes along some curve connecting two points in space–time. Since $\psi(\mathbf{r},\mathrm{t})$ satisfies the Schrödinger equation it must be continuous and therefore it is natural to assume that β must have a definite derivative almost at every point $(\mathbf{r},\mathrm{t})$. We will discuss later the points where this does not happen.

The change of β along a curve which does not pass through such singular points can be expressed by the integral $\sum_\mu \int \kappa_\mu dx_\mu$ taken along this curve with $\kappa_i = \partial\beta(\mathbf{r},\mathrm{t})/\partial x_i$ and $\kappa_0 = \partial\beta(\mathbf{r},\mathrm{t})/\partial t$. Since κ_μ in general do not satisfy the conditions of integrability $\partial\kappa_\mu/\partial x_\nu = \partial\kappa_\nu/\partial x_\mu$ the value of this integral depends on the curve and in particular the total change in the phase β need not vanish when the integral is calculated round a closed curve. The values of such circulation integrals for all imaginable closed curves completely characterize the multivalued properties of the non–integrable phase β.

We now show that in order to have unambiguous results for physical quantities *any such circulation integral must be the same for all the wave functions.* Indeed probabilities to measure physical quantities are given by squares of moduli of overlap integrals $\int \psi_m^* \psi_n d^3r$ with different wave functions ψ_m and ψ_n. In order that any such integral will have a definite modulus the integrand, although it need not have a definite phase at each point, must have a definite phase difference between any two points. Thus the change of phase of $\psi_m^*\psi_n$ round a closed curve must vanish. This

[9] In this section we draw freely on the original paper of P.A.M. Dirac, Ref. [15].

requires that the change in phase in ψ_n round a closed curve shall be equal to that in ψ_m and since ψ_m is arbitrary it must be a universal value for a given curve for all wave functions. This result means that without loss of generality the possible non integrable phase factor $\exp(i\beta)$ in the wave function may be taken as *universal* for all wave functions in a given physical problem.

Let us now consider the Schrödinger equation for ψ. Since

$$-i\hbar\frac{\partial}{\partial x}\psi = e^{i\beta}\left(-i\hbar\frac{\partial}{\partial x} + \hbar\kappa_x\right)\phi \tag{121}$$

with similar relations for the y, z and t derivatives one obtains that the single valued part ϕ of the general wave function ψ satisfies the Schrödinger equation with *gauge potentials* which are proportional to the derivatives of the non integrable phase β. In the most common case these would have to be identified with the electromagnetic potentials

$$\mathbf{A} = (\hbar c/e)\boldsymbol{\kappa} \ , \ A_0 = -(\hbar/e)\kappa_0 \ . \tag{122}$$

We therefore conclude that multivalued wave functions need not be considered in quantum mechanical description since they are equivalent to single valued wave functions in the presence of an external gauge field.

Although this conclusion is certainly valid there are two ambiguities which remain in the above discussion. The first is related to the Aharonov–Bohm effect and can occur in multiply connected regions such as the inside of a ring as was already discussed in Section 7 above. In this case even for a vanishing electromagnetic field *inside the region* one can not in general assume that the wave function must be single valued. If one can find non contractable closed curves in the region one must first classify these curves according to different homotopy classes as in Section 7. One may then assign an arbitrary but fixed phase factor $\exp(i\beta_k)$ for every elementary homotopy class C_k and demand that only solutions of the Schrödinger equation which change their phase by these assigned factors are allowed.

Intuitively one can interpret this situation by thinking about a multiply connected region as a region with "holes". Even when the electromagnetic field vanishes inside the region one can still have arbitrary magnetic fluxes "in the holes". These fluxes will give rise to Aharonov - Bohm phases for closed curves surrounding the "holes" provided these curves can not be continuously deformed to a point. Hence assigning different sets of phase factors $\exp(i\beta_k)$ for elementary classes of curves corresponds to assuming different distributions of Aharonov-Bohm fluxes $\exp(2\pi i\Phi_k/\Phi_0)$ in the "holes".

There exists another important ambiguity in the discussion of possible appearance of non integrable phases in quantum mechanics. This was first observed by Dirac and

is related to the fact that although *in the absence of the electromagnetic field in a singly connected region* the factor $\exp(i\beta)$ can be taken as single valued the phase β itself may change by an arbitrary integer multiple of 2π. Allowing for such changes requires a reconsideration of the connection between the derivatives κ of the non integrable phase β and the electromagnetic potentials and leads to a new physical phenomenon – a possible existence of *magnetic monopoles with quantized charges*. We will now discuss this fascinating subject.

11.2 Magnetic monopoles

The Maxwell equation $\nabla\cdot\mathbf{B} = 0$ means that there are no sources of the magnetic field, i.e. that the magnetic charges do not exist in nature. However nothing conceptually wrong should occur in the classical theory if one assumes a non zero $\nabla\cdot\mathbf{B} = 4\pi\,\sigma$ with σ – the density of magnetic charges. In fact the theory would be more symmetrical in this case since a symmetry under the so called duality transformation $\mathbf{E} \to \mathbf{B}$, $\mathbf{B} \to -\mathbf{E}$ would then exist if one simultaneously exchanges the magnetic and the electric charges. The non zero $\nabla \cdot \mathbf{B}$ poses however a problem in quantum theory where the canonical or path integral quantization in the presence of a magnetic field require an explicit introduction of the vector potential via $\mathbf{B} = \nabla \times \mathbf{A}$. Without this relation one is not able to define the Hamiltonian or the Lagrangian of the theory but it is valid only for divergenceless $\mathbf{B}$. Let us analyze this problem more closely and consider a hypothetical point–like particle, called magnetic monopole, which carries a magnetic charge g. In its presence

$$\nabla \cdot \mathbf{B} = 4\pi g\delta(\mathbf{r} - \mathbf{r}_0) \quad , \quad \mathbf{B} = g\frac{\mathbf{r} - \mathbf{r}_0}{|\mathbf{r} - \mathbf{r}_0|^3} \, , \tag{123}$$

where $\mathbf{r}_0$ denotes the position of the monopole and g is its magnetic charge.

For all points in space apart from an infinitesimal vicinity of $\mathbf{r}_0$ we have a divergenceless $\mathbf{B}$ and can write $\mathbf{B}(\mathbf{r}) = \nabla \times \mathbf{A}(\mathbf{r})$. Although correct *locally* the function $\mathbf{A}(\mathbf{r})$ is not single valued. This is seen by considering the integral form of the relation $\mathbf{B} = \nabla \times \mathbf{A}$, i.e. the Stokes theorem,

$$\int_S \mathbf{B} \cdot d\mathbf{S} = \oint_C \mathbf{A} \cdot d\mathbf{r}$$

where C is some closed curve in space and the integral on the left hand side is over an arbitrary surface S with C as its boundary. Such an integral – the flux of $\mathbf{B}$ – is however *not unique* in the present case. It does not change for all surfaces which

can be continuously deformed into each other without crossing the position of the monopole but once the surface crosses $\mathbf{r}_0$ the flux changes. The difference between the fluxes for surfaces "on both sides" of the monopole is equal to the total flux through the closed surface which these two surfaces form. Integrating (123) over the volume inside this surface and using the Gauss theorem one finds that this flux is equal to $4\pi\, g$. The non zero $\nabla \cdot \mathbf{B}$ thus effects the definition of $\mathbf{A}$ globally and not just near $\mathbf{r}_0$. Using the Stokes theorem with continuously changed contour C as a way of continuous definition of the relation between the functions $\mathbf{A}(\mathbf{r})$ and $\mathbf{B}(\mathbf{r})$ we will find two "branches" of this relation depending on "which side" of $\mathbf{r}_0$ we choose the surface S.

There is a number of ways of overcoming this difficulty. Historically the first was suggested by P. M. Dirac, Ref. [15]. We will follow a more modern way of presenting this approach. The idea is somewhat similar to what is done in the theory of multivalued analytic functions, i.e. to introduce a branch cut extending from a branch point. Viewing the position of the magnetic monopole as analogous to such a branch point one can avoid the ambiguity in the use of the Stokes theorem for determining the relation between $\mathbf{A}(\mathbf{r})$ and $\mathbf{B}(\mathbf{r})$ if together with $\mathbf{r}_0$ a thin tube extending from it to infinity (or to another, oppositely charged monopole) is excluded from the space. By excluding we mean that the surface S for the contour C can never be chosen such that it is pierced by the tube. This uniquely defines "the side" of the monopole which one should choose to draw the surface in the Stokes formula. One can thus assure the single valuedness of the $\mathbf{B}(\mathbf{r}) \to \mathbf{A}(\mathbf{r})$ relation everywhere in space apart from the inside of the excluded tube. We can make the tube as thin as we like and send it in any direction.

Let us illustrate this discussion and consider an example of a monopole placed at the origin and let us choose the excluded tube along a positive z axis. It is easy to verify that the vector potential the curl of which gives the magnetic field (123) everywhere except on the positive z can be chosen as

$$A_r = A_\theta = 0 \ , \ \ A_\phi = -\frac{g}{r}\frac{(1+\cos\theta)}{\sin\theta} \tag{124}$$

where A_r, A_θ and A_ϕ are spherical components of $\mathbf{A}$ and θ and ϕ are the conventional polar and azimuthal angles. On the z axis this potential does not reproduce the field (123) of the monopole but rather gives a singular magnetic field directed towards the monopole and carrying a flux $4\pi\, g$. The total effective magnetic field represented by the curl of (124) is therefore

$$\mathbf{B}_{eff} = g\frac{\mathbf{r}}{r^3} - g \ \ \delta(x)\,\delta(y)\,\mathbf{e}_z \ . \tag{125}$$

where $\theta(z)$ denote the step function.

We could choose another vector potential

$$A'_r = A'_\theta = 0 \ , \quad A'_\phi = \frac{g}{r}\frac{(1-\cos\theta)}{\sin\theta} \tag{126}$$

which also gives the required magnetic field (123) but with the excluded tube along the negative z axis. The corresponding "effective" field again has a singular component along this tube in the direction of the monopole. The flux along the tube is equal to the total flux of the first component of $\mathbf{B}_{eff}$, i.e. the flux of the monopole field.

It is easy to understand now the logic behind the excluded tube construction. The magnetic flux along the tube "feeds" the radially directed field of the monopole so that the resulting "effective" field is divergenceless, $\nabla \cdot \mathbf{B}_{eff} = 0$ and can be represented as a curl of a vector potential like the examples (124) and (126) above.

At this point a crucial question arises. We have replaced the desired magnetic field of the monopole by the effective field with the flux tube. How does one make sure that this modification has not changed the physics of the problem? Since the entire construction was invented for quantum mechanics we must worry if the presence of the flux tube added to the field of the monopole influences the solutions of the Schrödinger equation. In fact we know that such a flux tube does have a *global* influence in the form of the Aharonov–Bohm effect. It is also clear how to avoid this effect and make the flux tube *unobservable* at large distances. One must demand that the flux carried by the tube is equal to an integer multiple of the magnetic flux quanta,

$$4\pi\, g = n\Phi_0 = 2\pi n\hbar c/e$$

This imposes a *quantization condition* on the possible values of g,

$$eg = \frac{1}{2} n\hbar c \ , n = \pm 1, \pm 2, \ldots \tag{127}$$

This relation is called *the Dirac quantization condition*. The *unobservable flux tube* carrying integer number of magnetic flux quanta is called the *Dirac string*. The entire construction which we just described is called the *Dirac monopole*, cf., Fig. 7

The quantization condition (127) for the monopole charge implies that if there exists a magnetic monopole anywhere in the universe all electric charges will be quantized: $e = n(\hbar c/2g)$. Note that this quantization condition has an explicit dependence on the Plank constant and therefore on the quantum theory.

Experimental search for the presence of the magnetic monopoles in nature has so far given negative results. We note that since $g = (137/2)\, e$ the force between two

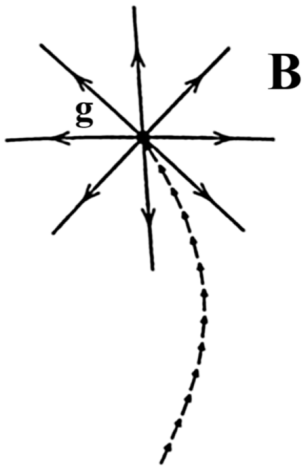

Figure 7: Magnetic field of Dirac monopole including the singular string. Note that the string is plotted as curved, which is allowed, but in the examples in the text it was chosen to be a straight line along the positive z axis for simplicity, Eq. (125).

monopoles is $(137/2)^2 \cong 4692$ larger than between two electrons. This may mean perhaps that all the monopoles in nature are tightly bound in pairs of opposite sign. In order to decide whether this is true one needs to know the masses of the monopoles about which the theory above gives no information.

In recent years another theory of magnetic monopoles was suggested by t 'Hooft and Polyakov, Ref. [17], in the context of the so called non abelian gauge theories with broken symmetry. This theory predicts that the mass of the monopoles should be very large. Viewed from large distances both Dirac and non abelian monopoles should look exactly the same and our discussion of quantum mechanics in the field of monopoles is expected to remain valid at such distances.

There is another way to introduce magnetic monopoles in quantum mechanics which avoids the appearance of the singular Dirac string. It was proposed by Wu and Yang, Ref. [16], and adopts an approach of sections similar to what is done by mapmakers when they map the spherical surface of the earth onto a plane map. A single map would obviously have a singularity at one point. Indeed imagine a rubber sheet with rectangular coordinate grid on it and try to wrap it around the globe. In order to avoid the singularity of a single map two maps are introduced, one for say a northern hemisphere and one for the southern. The two maps together form a singularity–free mapping of the earth. In order to be able to pass smoothly from

one map to another one should let each to cover more than its own hemisphere so that an overlap is created in the region of the equator. In this overlapping region the coordinates of both maps must be in one to one correspondence for identical points of the globe surface. In a similar way singularity-free vector potential can be found for a magnetic monopole.

The emerging formulation is in essence the so called *fiber bundle* formulation of gauge fields in quantum mechanics. We will not go into this here referring the reader to the literature.

11.3 Angular momentum and rotational symmetry in the presence of a monopole

Although the magnetic field of the monopole is spherically symmetric it should be intuitively clear that the Lorenz force acting perpendicular to the velocity of a moving particle will not conserve the ordinary expression $\mathbf{r} \times m\mathbf{v}$ for the angular momentum relative to the origin. Take, e.g., a particle which starts along a planar circular orbit around the monopole. The magnetic field will deflect it away from the plane changing the initial $\mathbf{r} \times m\mathbf{v}$. Using the equation of motion one can calculate the rate of change of this expression

$$\frac{d}{dt}(\mathbf{r} \times m\mathbf{v}) = \mathbf{r} \times m\frac{d}{dt}\mathbf{v} = \frac{eg}{cr^3}(\mathbf{r} \times (\mathbf{v} \times \mathbf{r})) = \frac{d}{dt}\left(\frac{eg}{c}\frac{\mathbf{r}}{r}\right) . \tag{128}$$

This suggests that we can define the total angular momentum as

$$\mathbf{J} = \mathbf{r} \times m\mathbf{v} - \left(\frac{eg}{c}\frac{\mathbf{r}}{r}\right) = \mathbf{r} \times \mathbf{p} - \frac{e}{c}\left[\mathbf{r} \times \mathbf{A} + g\frac{\mathbf{r}}{r}\right] , \tag{129}$$

so that it is conserved.

The appearance of unusual terms in the expressions of conserved quantities in the presence of electromagnetic field should already be familiar from the expressions of momenta in uniform electric and magnetic fields, Eqs. (47) and (76). In addition to the generators of the symmetry one must include the generator of the gauge transformation which is needed to keep the formulation in the "same" gauge. One can see this explicitly by considering infinitesimal rotation $\mathbf{r} \to \mathbf{r}' = \mathbf{r} + \delta\boldsymbol{\phi} \times \mathbf{r}$ and correspondingly

$$\mathbf{A}(\mathbf{r}) \to \mathbf{A}'(\mathbf{r}) = \mathbf{A}(\mathbf{r} - \delta\boldsymbol{\phi} \times \mathbf{r}) + \delta\boldsymbol{\phi} \times \mathbf{A}(\mathbf{r}) \to \mathbf{A}(\mathbf{r}) + \delta\boldsymbol{\phi} \times \mathbf{A}(\mathbf{r}) - [(\delta\boldsymbol{\phi} \times \mathbf{r}) \cdot \nabla]\, \mathbf{A}(\mathbf{r})$$

For $\mathbf{A}(\mathbf{r})$ which corresponds to the spherically symmetric magnetic field of the magnetic monopole the last term in this expression must be a gauge transformation, i.e.

equal to a gradient of a scalar function, $\nabla\xi(\mathbf{r})$. One finds

$$\xi(\mathbf{r}) = -\delta\boldsymbol{\phi}\cdot\left(\mathbf{r}\times\mathbf{A}(\mathbf{r}) + g\frac{\mathbf{r}}{r}\right) ,$$

The transformation of the wave function under rotation is therefore

$$[1+\frac{i}{\hbar}\delta\boldsymbol{\phi}\cdot(\mathbf{r}\times\mathbf{p})]\,[1+i\frac{e}{\hbar c}\xi(\mathbf{r})]\,\psi(\mathbf{r}) .$$

In the brackets of the expression for ξ one finds just the two terms which must be added (after multiplication by e/c) to the canonical $\mathbf{r}\times\mathbf{p}$ in order to obtain the conserved Eq.(129).

There exists another, quite different interpretation of the last term in the expression (129). It is the angular momentum contained in the electromagnetic field which exists in the space surrounding the moving particle and the fixed monopole. Using the expression for $\mathbf{B}$ of the monopole and $\mathbf{E} = e(\mathbf{r}-\mathbf{r}_0)/4\pi|\mathbf{r}-\mathbf{r}_0|^3$ for the electric field of the particle at $\mathbf{r}_0$ one indeed finds

$$\mathbf{L}_{em} = \int d^3r\ \mathbf{r}\times(\mathbf{E}\times\mathbf{B}) = \frac{egr_0}{4\pi}\int d^3r\ \frac{r^2(\hat{\mathbf{r}}\cos\theta - \hat{\mathbf{r}}_0)}{r^3(r^2+r_0^2-2\,r\,r_0\,\cos\theta)^{3/2}} \tag{130}$$

where θ is the angle between $\mathbf{r}$ and $\mathbf{r}_0$. Straightforward evaluation of the integral gives $-eg\,\mathbf{r}_0/|\mathbf{r}_0|$, which for $eg = \hbar c/2$ gives the last term in (129).

12 Non Abelian Gauge Fields

In Section 3.3 we discussed how the existence of the electromagnetic field could be predicted by demanding that a global symmetry of the free Schrödinger equation becomes local, i.e. by "gauging" this symmetry. We give now an example of what happens when a more complicated *non abelian* symmetry is gauged leading to a concept of a non abelian gauge field, Ref. [18]. Let us assume that particles in our theory in addition to spin carry another *discrete intrinsic dynamical* variable τ which we will tentatively call pseudospin and which may take two values, $\tau = 1, 2$. In analogy with the spin variables the wave functions will now carry an additional index τ so that $|\ \psi_\tau\ |^2$ integrated and summed over all other variables ($\mathbf{r}$, spin, etc.) gives the probability to measure this particular value of τ. We also have to introduce operators which act on the variables τ and in terms of which we shall represent all observable quantities involving this variable. These operators must be hermitian 2×2

matrices. One can write any such operator $O_{\alpha\beta}$ as a linear combination of a unit and Pauli matrices with real coefficients

$$O = O_0\delta_{\alpha\beta} + \sum_{a=1}^{3} O_a \tau^{(a)}_{\alpha\beta} \ , \tag{131}$$

$$\tau^{(1)} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} , \tau^{(2)} = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} , \tau^{(3)} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} ,$$

since the Pauli matrices together with a unit matrix represent a complete set for expanding any 2×2 matrix cf., Density Matrix chapter in the notes of this course.

Now let us consider a free Schrödinger equation

$$i\hbar\frac{\partial}{\partial t}\psi_\tau(\mathbf{r},\mathrm{t}) = -\frac{\hbar^2}{2m}\nabla^2\psi_\tau(\mathbf{r},\mathrm{t})$$

for a particle with the pseudospin. This equation is obviously invariant under a linear transformation $\psi\,'_\tau = \sum_{\tau'} S_{\tau\tau'}\psi_{\tau'}$ with an arbitrary matrix S. In order to preserve the normalization of ψ the matrix S must be unitary. Its general form therefore must be an imaginary exponential of an arbitrary 2×2 hermitian matrix and therefore can be written in terms of the Pauli matrices as

$$S = \exp\left(i\varphi_0 + i\sum_a \varphi_a \tau^{(a)}\right) \ . \tag{132}$$

with arbitrary *real* φ 's. For simplicity we will limit ourselves to the transformations with $\varphi_0 = 0$ which is equivalent to imposing the condition $\det S = 1$ on the allowed matrices S. We say that the Schrödinger equation is *invariant under the global SU(2) transformations* , i.e. under the transformations which belong to the group SU(2) of all unitary 2×2 matrices with unit determinant. This group is non abelian – two arbitrary SU(2) matrices in general do not commute. Let us now employ *The Gauge Principle* of Section 3.3 for this SU(2) symmetry and demand that our theory must be not only globally but also *locally* invariant under the above SU(2) transformations. This means that transformations with matrices S having their parameters ϕ_a as arbitrary functions of $\mathbf{r}$ and t should leave the Schrödinger equation invariant. The way to achieve the invariance under such *local gauge* transformations is to introduce the gauge field potentials which will compensate for the derivatives of S when the transformed

$$\psi\,'(\mathbf{r},\mathrm{t}) = S(\mathbf{r},\mathrm{t})\psi(\mathbf{r},\mathrm{t}) \tag{133}$$

is inserted in the Schrödinger equation. Since the derivatives of the matrix S are obviously also matrices the compensating potentials should be matrices. Since there are four derivatives $\partial/\partial t$ and $\nabla = (\partial/\partial x, \partial/\partial y, \partial/\partial z)$ in the Schrödinger equation one must introduce four such matrix compensating potentials $A^{(\mu)},\ \mu = 0, 1, 2, 3, 4$. They can be represented as linear combinations

$$A^{(\mu)}(\mathbf{r},\mathrm{t}) = \sum_{a=1}^{3} A_a^{(\mu)}(\mathbf{r},\mathrm{t})\frac{\tau^{(a)}}{2} = A^{(\mu)} \cdot \frac{\tau}{2}\ , \tag{134}$$

where we employed obvious short hand notation for the sum of products of arbitrary real functions $A_a^{(\mu)}(\mathbf{r},\mathrm{t})$ and Pauli matrices $\tau^{(a)}$ and introduced the factor $1/2$ to follow the conventional definitions in this field. In the fixed basis of $\tau^{(a)}$'s to represent a matrix $A^{(\mu)}$ is equivalent to giving three functions $A_a^{(\mu)}$.

In analogy with electromagnetism we introduce now gauge covariant derivatives

$$\begin{aligned} D_0 &= \frac{\partial}{\partial t} + \frac{ig}{\hbar}A^{(0)} \cdot \frac{\tau}{2}\ , \\ \mathbf{D} &= \nabla - \frac{ig}{\hbar c}\mathbf{A} \cdot \frac{\tau}{2}\ , \end{aligned} \tag{135}$$

and use them in the Schrödinger equation in place of the ordinary derivatives ,

$$i\hbar D_0\psi(\mathbf{r},\mathrm{t}) = -\frac{\hbar^2}{2m}\mathbf{D}^2\psi(\mathbf{r},\mathrm{t})\ . \tag{136}$$

The constant g introduced here is analogous to the electric charge e in the electromagnetism. It determines the strength of the coupling of the particle described by this equation to the non abelian gauge fields $A_a^{(\mu)}$. In order to achieve the invariance of the equation under the local gauge transformations (133) we demand that $D_0\psi$ and $\mathbf{D}\psi$ have the same transformation properties as ψ itself, i.e.

$$D_0\ '\psi\ ' \equiv \left(\frac{\partial}{\partial t} + \frac{ig}{\hbar}A^{(0)\ \prime} \cdot \frac{\tau}{2}\right)\psi\ \ ' \ = \ SD_0\psi \equiv S\left(\frac{\partial}{\partial t} + \frac{ig}{\hbar}A^{(0)} \cdot \frac{\tau}{2}\right)\psi \tag{137}$$

$$\mathbf{D}\ \ '\psi\ ' \equiv \left(\nabla - \frac{ig}{\hbar c}\mathbf{A}\ ' \cdot \frac{\tau}{2}\right)\psi\ ' \ = \ S\mathbf{D}\psi \equiv S\left(\nabla - \frac{ig}{\hbar c}\mathbf{A} \cdot \frac{\tau}{2}\right)\psi\ . \tag{138}$$

This is obviously a sufficient condition for the invariance of Eq.(136) and determines the transformation properties of the gauge potentials

$$A^{(\mu)\prime}(\mathbf{r},\mathrm{t}) \cdot \frac{\tau}{2} = S(\mathbf{r},\mathrm{t})A^{(\mu)}(\mathbf{r},\mathrm{t}) \cdot \frac{\tau}{2}S^{-1}(\mathbf{r},\mathrm{t}) - \frac{i}{g}\left[\frac{\partial S(\mathbf{r},\mathrm{t})}{\partial x_{(\mu)}}S^{-1}(\mathbf{r},\mathrm{t})\right]\ . \tag{139}$$

We see that under a gauge transformation each matrix gauge potential is locally "rotated" at every space–time point by the gauge transformation matrix $S(\mathbf{r},t)$ and at the same time it is shifted by an amount which depends on the corresponding derivative of $S(\mathbf{r},t)$. This expression as well as the relations above are valid for any unitary group SU(N) with the appropriate generalization of the transformation matrix S and the Pauli matrices $\tau^{(a)}$. For the abelian group U(1) we will obviously recover the known Schrödinger equation and the gauge potentials of the electromagnetic field. In general there will be $d \times (N^2-1)$ gauge potentials with $d = 4$ – the dimensionality of the space–time and (N^2-1) – the number of the independent generators of the group SU(N). The gauge freedom expressed by (139) means that in general only $(d-1) \times (N^2-1)$ combinations of the gauge potentials are independent. We finally remark that among the gauge fields known in nature the unified electromagnetic and the weak interactions (often called electroweak) are described by $U(1) \times SU(2)$ and the strong interactions by SU(3) non abelian gauge potentials. The intrinsic quantum numbers for these interactions (analog of what we called pseudospin) are the standard electric charge and the so called weak isospin and the color respectively. Since as already mentioned the gravitational field is also a gauge field we have all four basic interactions described by the gauge fields.

Problems

1. Consider the spin part of the time reversal operator $U = \exp[i2\pi\, s_y/\hbar]$.

   - Show that U = -1 when applied to the wave function of a spin 1/2 particle.
   - Components ψ_{jm} of a spin wave function with a general spin j can be considered as far as their transformation properties are concerned as suitably chosen components of the wave function of a system of 2j spin 1/2 particles. Use this to prove that $U\psi_{jm} = (-1)^{2j}\psi_{jm}$.

2. Find the propagator in a uniform electric and magnetic fields by evaluating the appropriate Gaussian path integrals, cf., [5]

3. Electrons are confined to move in a plane (x, y) and are placed in a uniform magnetic field perpendicular to the plane.

   (a) Consider two different gauges choices **a)** $A_x = -By$, $A_y = A_z = 0$ and **b)** $A_x = -\frac{1}{2}By$, $Ay = \frac{1}{2}Bx$, $A_z = 0$. electron eigenfunctions calculated in Which combinations of the guiding center coordinates do they diagonalize?

How are these two sets of eigenfunctions related? Calculate and explain the behaviour of the current density in each of the above cases. What is the total current? How the current will change in the case **a)** above if one adds a uniform electric field along the x direction? along the y direction? What is the total current now? In which direction does it flow? In case **b)** above assume that a very thin solenoid with magnetic flux Φ is added along the z axis (at $x = y = 0$). What and how will it influence? Consider your answer for various values of the Φ and see if there are special values of ϕ.

In addition they are su a potential $U = \frac{1}{2}\alpha x^2$.

a) How the Landau levels are changed by this potential? What is the current density in a single state of a Landau level as compared to the case with $U = 0$?

b) Now repeat for $U = \frac{1}{2}\alpha(x^2 + y^2)$.

2. Show that the operators x_o and y_o of the guiding center coordinates are generators of the translations in the presence of the uniform magnetic field.

3. Consider wave functions of Landau levels with definite values of x_o. How are they related to the similar wave functions with y_o? Consider now wave functions with definite $x_o^2 + y_0^2$. Assume that a very thin solenoid with magnetic flux ϕ?

4. Consider quantum mechanics in a strong magnetic field.

   a.The projection on a lowest Landau level (LLL). Show that the eigenstates of x_0 and y_0 are exactly equivalent to eigenstates of p and x in a one dimensional quantum mechanics.

   b.Derive semiclassical approximation in the limit of $\ell \to 0$.

   c. Find semiclassical energy levels of two interacting electrons in 2 dimensions in a strong magnetic field. Discuss also the case of oppositely charged particles (say an electron - positron system or electron-hole system in a solid state).

“

References and Guide to Reading.

[1] D. J. Griffith, Introduction to Electrodynamics, 4th edition, Cambridge University Press.

[2] Ta-Pei Cheng and Ling-Fong Li, Gauge theory of elementary particle physics, Clarendon Press, Oxford, 1984. This and the following reference are fairly advanced quantum field theory courses which however contain several chapters which could be read with a profit even by a less advanced student.

[3] Lewis H. Ryder, Quantum Field Theory, Cambridge University Press, 1985.

[4] L. S. Schulman, Techniques and Applications of Path Integration, Wiley, New York, 1981. Section 4 and 5 contain a pedagogical discussion of the subtleties of path integration with vector potentials and of the related "Ito Integral". Another version of the derivation is found in L.S. Schulman, Selected Topics in Path Integration, in Lectures on Path Integration: Trieste 1991, H.A. Cerdeira et al, Eds.,World Scientific, Singapore.

[5] Hagen Kleinert, Path Integrals in Quantum Mechanics, Statistics, and Polymer Physics, World Scientific, Singapore, 1989. This volume contains several sections with discussions of path integrals in the presence of magnetic field.

[6] L. Landau and E. Lifshitz, Quantum Mechanics, Pergamon Press, 1965

[7] Y. Aharonov and D. Bohm, Phys. Rev., 115 (1959) 484 – this is the original publication in which phenomena later called Aharonov–Bohm effect were described. The effect was first experimentally seen by R. G. Chambers, Phys. Rev. Lett. 5 (1060) 3. A very lucid discussion of it appears in T. T. Wu and C. N. Yang, Phys. Rev. D12 (1975) 3845.

[8] Our discussion of the Aharonov-Bohm version of the two slit experiment is based on R. P. Feynman, R. B. Leighton and N. Sands, The Feynman Lectures on Physics, vol.2, Sec. 15.5, Reading, Massachusetts: Addison-Wesley Publishing, 1964.

[9] Palash B. Pal, A Physicist's Introduction to Algebraic Structures, Ch. 22, Cambridge University Press, 2019.

[10] F. Bloch, Phys. Rev., 2 (1970) 109. Extremely clear and beautifully argumented "old master's" paper exploring in the context of the Josephson effect in a superconducting ring the most general consequences of quantum mechanics in multiply connected regions in the presence of Aharonov–Bohm fluxes.

[11] Properties of the Airy function can be found in, e.g. M. Abramovitz and I. A. Stegun, Handbook of Mathematical Functions. Washington, DC: National Bureau of Standards.

[12] J. J. Sakurai, Modern Quantum Mechanics (Revised Edition) , Addison-Wesley, p. 239

[13] Ho-Fai Cheung et al, Phys. Rev. B37 (1988) 6050

[14] B. I. Halperin, Phys. Rev. B25 (1982) 2185

[15] The first paper on magnetic monopoles, P. A. M. Dirac, Proc. Roy. Soc. A133 (1931) 60 was in fact an investigation of a more general subject of justification of the demand that wave functions are single valued. It is strikingly modern, highly readable and is reprinted in a more accessible Monopoles in Quantum Field Theory, Proceedings of the Monopole Meeting, Trieste, Italy, December 1981, N. S. Craigie, P Goddard and W. Nahm, Eds., World Scientific. The continuation of the first paper followed in P. A. M. Dirac, Phys. Rev. 74 (1948) 817.

[16] T.T. Wu and C.N. Yang, Nucl. Phys., B107 (1976) 365

[17] G. tH́ooft, Nucl. Phys., B79 (1974) 276; A.M. Polyakov, JETP Letters, 20 (1974) 194.

[18] C. N. Yang and R. L. Mills, Phys. Rev., 96 (1954) 191. This is the original paper in which the non abelian generalization of the electromagnetism, later called the Yang–Mills field, was first proposed. In Chapter 3.5 of Ref. [3] a pedagogical review of the Yang–Mills fields is presented.