MONASH
UNIVERSITY

QUANTUM MECHANICS
FOURTH YEAR LECTURE NOTES
David Paganin, 2004

David Paganin
Room 159, Building 19
Email: David.Paganin@spme.monash.edu.au
Ph: 9905 3633

CHAPTER 1

RELATIVISTIC QUANTUM MECHANICS AND THE DIRAC EQUATION

§1. Introduction and overview

In this first chapter of the course, we shall study relativistic quantum mechanics, which - as the title implies - is a formulation of the principles of quantum mechanics which is consistent with Einstein's Special Theory of Relativity. This takes us a step beyond the (non-relativistic) Schrödinger wave mechanics which you will have encountered in previous studies.

The remainder of this section is devoted to a brief overview of the rest of the chapter. My aim in giving you a summary up-front is to help you get the "big picture" as early as possible, aiding you in your studies of this difficult material.

Section §2 introduces the so-called "natural units" which are often used in theoretical physics. In these units, the reduced Planck constant $\hbar \equiv \mathrm{h}/(2\pi)$ and the speed of light c have a value of unity. We shall be using these units for the remainder of this chapter, with the exception of section §5.6.

Section §3 is broken up into two parts. Section §3.1 revises your previously-acquired knowledge regarding the time-dependent Schrödinger equation, emphasizing the fact that this equation does not take into account the relativistic modifications to Newtonian mechanics which are accounted for in Einstein's theory of Special Relativity. In Section §3.2, we make a first attempt at formulating a relativistically-correct form of the Schrödinger equation; the equation we so derive will be seen to be unsatisfactory, for reasons which we shall go into in some detail.

Section §4 is devoted to the relativistic equation known as the Klein-Gordon equation. In §4.1 we point out that, historically, this equation was derived in an attempt to overcome the problems described in §3.2. In §4.2, we give a derivation of the Klein-Gordon equation in a manner which explicitly takes Special Relativity into account. We then explore this equation in §4.3, by deriving expressions for the current and probability density of the Klein-Gordon wavefunction. At this point, we again encounter difficulties which lead us to discard the Klein-Gordon equation.

Section §5 deals with the famous Dirac equation, which is a relativistic quantum-mechanical equation for an electron in an electromagnetic field. Section §5.1 introduces a mathematical identity which will be useful in our discussions of the Dirac equation. In §5.2 we derive the Dirac equation, for the case of free space (where no electromagnetic fields are present to "push the electron around"). We then explore this equation in §5.3, by deriving expressions for the current and probability density of the Dirac wavefunction. At this point, we see that the difficulties which led us to discard the Klein-Gordon equation, are not applicable to the Dirac equation. At last, we shall have found a satisfactory relativistic quantum-mechanical description for the electron. So far, we have been working in free space, but in §5.4 we consider the modifications which must be made to the Dirac equation in order to describe the behaviour of the electron in the presence of an electromagnetic field. In §5.5, we consider the non-relativistic limit of this Dirac equation for an electron in an electromagnetic field, and find that we obtain - as we must! - the correct equation from non-relativistic quantum mechanics (the so-called Schrödinger-Pauli equation for a non-relativistic electron in the presence of an electromagnetic field; this equation reduces to the Schrödinger equation when the electromagnetic field is "turned off"). Section §5.6 uses the Dirac equation to obtain the first-order relativistic corrections to the Schrödinger equation for an electron in the presence of an electrostatic potential; we shall see that the spin-orbit force, together with other terms known as the Thomas and Darwin terms, are automatic consequences of the Dirac theory; this may be contrasted with the relatively *ad hoc* introduction of these terms in the context of non-relativistic quantum mechanics. These additional terms may be used to determine the first-order relativistic corrections to the energy levels of the hydrogen atom.

§2. Note on units

In this chapter on relativistic quantum mechanics, we shall work with the so-called "natural units"

where both the reduced Planck constant $\hbar \equiv h/(2\pi)$ and the speed of light c have a value of unity:

(1) $\hbar = c = 1$.

We make one exception to this chapter's use of natural units, in §5.6. Our reason for deviating from natural units in this particular section is that we shall need to keep track of various powers of c, which is a little difficult when c is set to 1!

§3. Non-relativistic and relativistic forms of the Schrödinger equation

This section is broken into two parts. We begin with §3.1, which re-acquaints you with the (non-relativistic) time-dependent Schrödinger equation. We emphasise that this equation is derived from considerations based on Newtonian mechanics, and is therefore inconsistent with Einstein's Special Relativity. This leads us to section §3.2, which is a first attempt at formulating a relativistic version of the Schrödinger equation. This attempt is ultimately abandoned, for reasons we shall discuss.

§3.1 Non-relativistic Schrödinger equation[1]

We can "derive" the (non-relativistic) Schrödinger equation by first considering the Newtonian expression for the energy E of a non-relativistic free particle of mass m, velocity $\vec{v}$ and momentum $\vec{p} = m\vec{v}$:

(2) $E = \frac{1}{2}m|\vec{v}|^2 = \frac{|m\vec{v}|^2}{2m} = \frac{|\vec{p}|^2}{2m}$.

Post-multiply[2] both sides of this equation by the time-dependent wavefunction $\Psi(\vec{r},t)$:

(3) $E\Psi(\vec{r},t) = \frac{|\vec{p}|^2}{2m}\Psi(\vec{r},t)$,

where $\vec{r}$ denotes a position vector in three-dimensional space and t denotes time. If we make use of the usual substitution ("Schrödinger correspondence rules"[3]):

(4) $E \to i\partial_t$, $\vec{p} \to -i\nabla$,

where $\partial_t \equiv \partial/\partial t$[4], then (2) becomes the familiar time-dependent Schrödinger equation:

(5) $i\partial_t \Psi(\vec{r},t) = \frac{-\nabla^2}{2m}\Psi(\vec{r},t)$.

§3.2 Relativistic Schrödinger equation[5]

The time-dependent Schrödinger equation (5) is clearly incompatible with Einstein's Special Relativity, because it is based on the non-relativistic Newtonian relation (2) between energy and momentum. Having made this observation, it is natural for us to see what equation we end up with if we apply the Schrödinger correspondence rules (4) to a *relativistic* relation between energy and momentum. To this end, we use the following result from Special Relativity, which states that the energy E has contributions from both the rest mass m and the momentum $\vec{p}$:[6]

(6) $E^2 = |\vec{p}|^2 + m^2$.

(Note that, if we had not been using natural units, this equation would read $E^2 = |\vec{p}|^2c^2 + m^2c^4$, reducing to the famous expression $E = mc^2$ when the momentum is zero.) Equipped with (6), we can now write down a relativistic version of (2)[7]:

(7) $E = \sqrt{|\vec{p}|^2 + m^2}$,

an equation which can be post-multiplied by the wavefunction to get a relativistic version of (3):

---

[1] R. H. Landau, *Quantum Mechanics II: A Second Course in Quantum Theory*, John Wiley & Sons, New York (1996), section 13.1.

[2] If we post-multiply the equation A=B by C, we get AC=BC. For later reference, note that if we were to pre-multiply the equation A=B by C, we get CA=CB.

[3] These are covered in elementary courses on quantum mechanics, so we will not revise them here. See, for example, B. H. Bransden and C. J. Jochain, *Introduction to Quantum Mechanics*, Longman Scientific and Technical, Essex (1989), section 3.1.

[4] Note that "≡" means "is by definition equal to".

[5] R. H. Landau, *Quantum Mechanics II: A Second Course in Quantum Theory*, John Wiley & Sons, New York (1996), section 13.2; J. D. Bjorken and S. D. Drell, *Relativistic Quantum Mechanics*, McGraw-Hill Book Company, New York (1964), pp. 4-5.

[6] A derivation of this equation is given in most textbooks on Special Relativity. See, for example, A. P. French, *Special Relativity*, Chapman and Hall, London (1968), pp. 21-22.

[7] You might like to show that (7) reduces to (2), up to an additive constant, for sufficiently-small momentum.

(8) $E\Psi(\vec{r},t)=\sqrt{|\vec{p}|^2+m^2}\,\Psi(\vec{r},t)$.

We note in passing that the negative root in (7) has been discarded as unphysical. Making use of the Schrödinger correspondence rules (4), we obtain the so-called *relativistic Schrödinger equation*:

(9) $i\partial_t\Psi(\vec{r},t)=\sqrt{-\nabla^2+m^2}\,\Psi(\vec{r},t)$.

There are a number of reasons for rejecting this equation. (i) We are faced with the mathematical problem of interpreting the rather bizarre-looking square-root operator on the right side of (9)[8]. If we expand this square root as an infinite power series then we will have an equation containing all powers of the derivative operator ∇^2, which is difficult to handle. (ii) Leading on from this comment, we see that spatial derivatives and time derivatives appear in this equation in a very non-symmetric form, which is difficult to reconcile with the notion of Special Relativity which treats space and time on a similar footing. (iii) As discussed in Appendix A (which is not examinable), equation (9) exhibits the distasteful feature of non-locality[9]. If interested, please work through Appendix A, which examines this point in more detail.

In view of the objections which have been raised against (9), we will reject it and try a different approach to the question of formulating a relativistic quantum mechanics.

§4. The Klein-Gordon equation[10]

§4.1 Introduction and overview

In section §3.2 we made an attempt to derive a relativistic form of the Schrödinger equation, an effort which culminated in equation (9). We deemed this equation to be unsatisfactory for a number of reasons, including (i) Difficulties in interpreting the square-root operator; (ii) an obvious lack of symmetry between the treatment of space and time coordinates which masks the relativistic nature of the equation; (iii) the non-local nature of the equation.

In the historical development of the quantum theory, an attempt was made to deal with these difficulties via another relativistic equation which is now known as the *Klein-Gordon equation*. We devote this section to the derivation and study of this equation.

§4.2 Derivation of the Klein-Gordon equation

Take the relativistic relation (6) between energy and momentum, apply the Schrödinger correspondence rules (4):

(10) $(i\partial_t)^2=(-i\nabla)^2+m^2$

and then post-multiply the resulting expression by the wavefunction $\Psi(\vec{r},t)$:

(11) $-\partial_t^2\Psi(\vec{r},t)=(-\nabla^2+m^2)\Psi(\vec{r},t)$.

This may be re-written as:

(12) $(\partial_t^2-\nabla^2+m^2)\Psi(\vec{r},t)=0$.

If we now introduce the D'Alembertian operator:

(13) $\Box\equiv\partial_t^2-\nabla^2$,

then (12) may be written in the compact form:

(14) $(\Box+m^2)\Psi(\vec{r},t)=0$.

Equation (14) is the famous *Klein-Gordon equation*.

Difficulty #1: The Klein-Gordon equation (14) is a *second-order* differential equation with respect to time, i.e. it contains the second derivative with respect to time. This may be contrasted with time-dependent Schrödinger equation (5), which is a first-order differential equation with respect to time. This difference between "first order in time" and "second order in time" is in fact rather profound. To explain this point further, let us compare the equation for the diffusion of heat for a potato placed in hot water (this equation is first order in time, like the time-dependent Schrödinger equation) to the equation for a vibrating string (which is

[8] Certain methods of modern mathematics, namely the theory of Fourier integral operators, allow such operators to be handled in a meaningful way. This is beyond the scope of the course, but I would be happy to explain this point in more detail to anyone who is interested!

[9] In this context, "non-local" means that the time-evolution of the wavefunction at a given point in space and at some given instant of time, is governed by the value of the wavefunction through all space, at that instant of time (see the final equation of Appendix A). Thus it appears that physically-separated parts of the wavefunction are able to instantaneously influence one another.

[10] A. Messiah, *Quantum Mechanics (volume II)*, North-Holland Publishing Company, Amsterdam (1961), pp. 884-888.

second order in time, like the Klein-Gordon equation) [11]. When we solve the problem of determining the time-dependent temperature distribution T(x,y,z,t) of a potato in hot water, we need only specify the "initial condition" of the temperature distribution of the potato and water at some instant of time t=0, namely T(x,y,z,t=0). For the case of determining the time-dependent displacement D(x,y,z,t) of the vibrating string, however, it is not sufficient to specify the initial displacement D(x,y,z,t=0) of the string at t=0; we also need to specify how *fast* the string is moving at t=0, i.e. $(\partial/\partial t)D(x,y,z,t=0)$. Similarly, to solve for the time-dependent wavefunction $\Psi(x,y,z,t)$ obeying the Schrödinger equation, we need only specify the wavefunction at some initial instant of time t=0, namely $\Psi(x,y,z,t=0)$; to solve for the time-dependent wavefunction $\Psi(x,y,z,t)$ obeying the Klein-Gordon equation, we must specify both $\Psi(x,y,z,t=0)$ and $(\partial/\partial t)\Psi(x,y,z,t=0)$. We would prefer to have a relativistic wave equation which is first-order in time, so that the value of the wavefunction at a given instant of time shall provide sufficient information to determine its evolution through time.

Difficulty #2: In working with (6) rather than (7), we have introduced an *extraneous negative energy root*:

$$(15)\quad E = -\sqrt{|\vec{p}|^2 + m^2}$$

in addition to the positive-energy root (7). We therefore seem to have sacrificed positive energy in our quest for a simple relativistic equation which does not suffer from the difficulties of (9)[12].

►*Exercise 1: Show that the following wavefunctions:*

$$(16)\quad \Psi(\vec{r},t) = \exp(imt), \quad \Psi(\vec{r},t) = \exp(-imt)$$

are negative and positive solutions to the Klein-Gordon equation, respectively. Interpret your results.

[11] The analogy of the potato and the string is taken from J. J. Sakurai, *Advanced Quantum Mechanics*, Addison-Wesley, Reading (1967), p. 91.

[12] This perceived difficulty may be surmounted by interpreting the negative-energy solutions as antiparticles.

§4.3 Current and probability density

Our next task considers the interpretation of the Klein-Gordon wavefunction $\Psi(\vec{r},t)$ which appears in (14). To achieve this, we shall derive expressions for the probability density P and probability current $\vec{j}$ satisfying the continuity equation:

$$(17)\quad \partial_t P + \nabla \bullet \vec{j} = 0 .$$

To this end, form two equations, the first of which is (14) pre-multiplied by $\Psi^*(\vec{r},t)$:

$$(18a)\quad \Psi^*(\vec{r},t)(\Box + m^2)\Psi(\vec{r},t) = 0 ,$$

and the second of which is the complex conjugate of (14) pre-multiplied by $\Psi(\vec{r},t)$:

$$(18b)\quad \Psi(\vec{r},t)(\Box + m^2)\Psi^*(\vec{r},t) = 0 .$$

If we subtract these two equations, we get:

$$(19)\quad \Psi^*\Box\Psi - \Psi\Box\Psi^* = 0 ,$$

where we have dropped the explicit functional dependence of the wavefunction on position and time (so as to avoid too much "clutter" in our equations!). Make use of the definition (13) of the D'Alembertian in (19), and then separate the terms involving time derivatives from those involving space derivatives, to arrive at:

$$(20)\quad \Psi^*\partial_t^2\Psi - \Psi\partial_t^2\Psi^* + \Psi\nabla^2\Psi^* - \Psi^*\nabla^2\Psi = 0 .$$

We leave it as an exercise for the reader to show that this may be re-written in the form:

$$(21)\quad \partial_t(\Psi^*\partial_t\Psi - \Psi\partial_t\Psi^*) + \nabla \bullet (\Psi\nabla\Psi^* - \Psi^*\nabla\Psi) = 0 .$$

►*Exercise 2: Derive the identity:*

$$(22)\quad \nabla \bullet (A\nabla B) = \nabla A \bullet \nabla B + A\nabla^2 B ,$$

and then make use of it to show that equation (20) leads to equation (21). Explain the reasoning which led us to perform these manipulations.

Comparison of (21) with (17) allows us to write down the following expressions for the probability density P and the probability current $\vec{j}$ in terms of the Klein-Gordon wavefunction $\Psi(\vec{r},t)$:

(23a) $\mathrm{P} = \Theta\left(\Psi^*\partial_t\Psi - \Psi\partial_t\Psi^*\right)$,

(23b) $\vec{\mathrm{j}} = \Theta\left(\Psi\nabla\Psi^* - \Psi^*\nabla\Psi\right)$,

where Θ is a constant which remains to be determined.

In order to find out the value of the constant Θ, we demand that the non-relativistic limit of (23a) shall give the familiar expression[13]:

(24) $\mathrm{P} \to \Psi^*\Psi$.

To this end, we leave it as an exercise to show that substitution of the *ansatz*[14]:

(25) $\Psi \equiv \widetilde{\Psi}\mathrm{e}^{-\mathrm{imt}}$,

into (23a) leads to:

(26) $\mathrm{P} = \Theta\left(-2\mathrm{im}\left|\widetilde{\Psi}\right|^2 + \widetilde{\Psi}^*\partial_t\widetilde{\Psi} - \widetilde{\Psi}\partial_t\widetilde{\Psi}^*\right)$.

►*Exercise 3: Is there any assumption involved in writing down the ansatz (25)? Use this ansatz to derive equation (26). Why do you think that (26) might be helpful in the context of our search for the constant Θ ?*

In the non-relativistic limit, the first term of (26) dominates because the rest-mass energy m is much larger than the particle kinetic energy $\mathrm{i}\partial_t$. Hence, in the non-relativistic limit, (26) becomes:

(27) $\mathrm{P} \to -2\mathrm{im}\Theta\left|\widetilde{\Psi}\right|^2 = -2\mathrm{im}\Theta\left|\Psi\right|^2$,

where the last step follows from taking the modulus squared of equation (25). Since we require (27) to conform to (24), we see that:

(28) $-2\mathrm{im}\Theta = 1 \quad\Rightarrow\quad \Theta = \dfrac{1}{-2\mathrm{im}} = \dfrac{\mathrm{i}}{2\mathrm{m}}$.

Thus we have achieved our aim of using our knowledge of the non-relativistic limit of (23a) to derive an expression for the constant Θ. Now that we know the value of this constant, we see that equations (23) become:

(29a) $\mathrm{P} = \dfrac{\mathrm{i}}{2\mathrm{m}}\left(\Psi^*\partial_t\Psi - \Psi\partial_t\Psi^*\right)$,

(29b) $\vec{\mathrm{j}} = \dfrac{\mathrm{i}}{2\mathrm{m}}\left(\Psi\nabla\Psi^* - \Psi^*\nabla\Psi\right)$.

These are the desired expressions for the probability density P and probability current $\vec{\mathrm{j}}$ in terms of the Klein-Gordon wavefunction $\Psi(\vec{\mathrm{r}},\mathrm{t})$.

From equations (26) and (28), we have:

(30)

$$\begin{aligned}
\mathrm{P} &= \frac{\mathrm{i}}{2\mathrm{m}}\left(-2\mathrm{im}\left|\widetilde{\Psi}\right|^2 + \widetilde{\Psi}^*\partial_t\widetilde{\Psi} - \widetilde{\Psi}\partial_t\widetilde{\Psi}^*\right) \\
&= \left|\widetilde{\Psi}\right|^2 + \frac{\mathrm{i}}{2\mathrm{m}}\left(\widetilde{\Psi}^*\partial_t\widetilde{\Psi} - \widetilde{\Psi}\partial_t\widetilde{\Psi}^*\right) \text{... isolate first term} \\
&= \left|\widetilde{\Psi}\right|^2 + \frac{1}{2\mathrm{m}}\left(\mathrm{i}\widetilde{\Psi}^*\partial_t\widetilde{\Psi} - \mathrm{i}\widetilde{\Psi}\partial_t\widetilde{\Psi}^*\right) \text{... bring "i" in brackets} \\
&= \left|\widetilde{\Psi}\right|^2 + \frac{1}{2\mathrm{m}}\left(\mathrm{i}\widetilde{\Psi}^*\partial_t\widetilde{\Psi} + \left(\mathrm{i}\widetilde{\Psi}^*\partial_t\widetilde{\Psi}\right)^*\right) \\
&= \left|\widetilde{\Psi}\right|^2 + \frac{1}{\mathrm{m}}\mathrm{Re}\left(\mathrm{i}\widetilde{\Psi}^*\partial_t\widetilde{\Psi}\right)
\end{aligned}$$

Difficulty #3: Because of the presence of the second term in equation (30), we see that the probability density associated with the Klein-Gordon equation is not necessarily positive.

Historically, the three noted "difficulties" led to a discarding of the Klein-Gordon equation. We too shall use these "difficulties" as an excuse to discard the Klein-Gordon equation. Note, however, that more advanced treatments[15] show that the Klein-Gordon equation can in fact provide a useful relativistic quantum theory for spinless particles such as π-mesons and K-mesons. This more sophisticated treatment involves an interpretation of the negative-energy solutions as antiparticles, and P as a charge density rather than a probability density (a charge density can happily be either positive or negative, unlike a probability density which must always be non-negative).

§5. The Dirac equation

Even though the Klein-Gordon equation is perfectly satisfactory when properly interpreted, there is still reason for rejecting it in the context of the

[13] Our "new" theory of relativistic quantum mechanics must agree with the old in the domain of validity of the latter.

[14] An *ansatz* is a mathematical assumption, especially about the form of an unknown function, which is made in order to facilitate solution of an equation or other problem.

[15] See, for example, Chapter 3 of F. Mandl and G. Shaw, *Quantum Field Theory (Revised Edition)*, John Wiley and Sons, Chichester (1993).

relativistic quantum mechanics of the *electron*. In particular, it cannot accommodate the spin-half nature of the electron as naturally as the so-called Dirac equation can. This section is devoted to a study of the Dirac equation, which is a *relativistic quantum-mechanical equation for spin-half particles such as the electron*.

In view of the length of this section, we break it up into a number of sub-sections.

§5.1 A mathematical preliminary

In this section we introduce the following extremely useful identity:

(31) $\left(\vec{\sigma}\bullet\vec{A}\right)\left(\vec{\sigma}\bullet\vec{B}\right)=\vec{A}\bullet\vec{B}+i\vec{\sigma}\bullet\left(\vec{A}\times\vec{B}\right)$,

where $\vec{A}$ and $\vec{B}$ are two vectors (or vector operators whose components commute with those of $\vec{\sigma}$), and $\vec{\sigma}$ is the matrix vector operator of Pauli matrices:

(32) $\vec{\sigma}\equiv\left(\sigma_x,\sigma_y,\sigma_z\right)$,

(33) $\sigma_x=\begin{pmatrix}0&1\\1&0\end{pmatrix},\ \sigma_y=\begin{pmatrix}0&-i\\i&0\end{pmatrix},\ \sigma_z=\begin{pmatrix}1&0\\0&-1\end{pmatrix}$.

You should have encountered the Pauli matrices in your third-year studies on spin in quantum mechanics. (If you have forgotten this material, I recommend that you revise your knowledge of the Pauli matrices and the two-component Pauli wavefunction by looking up an elementary quantum-mechanics text. I will provide handouts if requested!) The Pauli matrices satisfy the algebra:

(34a) $\sigma_x^2=\sigma_y^2=\sigma_z^2=1$,

(34b) $\sigma_x\sigma_y=-\sigma_y\sigma_x=i\sigma_z$,

(34c) $\sigma_y\sigma_z=-\sigma_z\sigma_y=i\sigma_x$,

(34d) $\sigma_z\sigma_x=-\sigma_x\sigma_z=i\sigma_y$.

►*Exercise 4: Show that the Pauli matrices (33) indeed satisfy the algebra in equations (34).*

► *Exercise 5: Prove equation (31). Hint: You might want to start the proof by first writing the vector of Pauli matrices component-wise as in equation (32), and writing the vector* $\vec{A}=\left(A_x,A_y,A_z\right)$*, and similarly for* $\vec{B}$.

§5.2 Derivation of the Dirac equation[16]

Take equation (31), setting both $\vec{A}$ and $\vec{B}$ equal to the momentum operator $\vec{p}$, leading to:

(35) $\left(\vec{\sigma}\bullet\vec{p}\right)\left(\vec{\sigma}\bullet\vec{p}\right)=\vec{p}\bullet\vec{p}+i\vec{\sigma}\bullet\left(\vec{p}\times\vec{p}\right)$.

Since $\vec{p}\times\vec{p}=0$[17], we arrive at:

(36) $\left(\vec{\sigma}\bullet\vec{p}\right)\left(\vec{\sigma}\bullet\vec{p}\right)=\vec{p}\bullet\vec{p}$.

Thus the relativistic energy-momentum relation in equation (6) can be re-written as:

(37) $E^2-\left(\vec{\sigma}\bullet\vec{p}\right)\left(\vec{\sigma}\bullet\vec{p}\right)=m^2$.

The left-hand side of this operator equation can be factorised as a difference of perfect squares[18]:

(38) $\left(E-\vec{\sigma}\bullet\vec{p}\right)\left(E+\vec{\sigma}\bullet\vec{p}\right)=m^2$.

Post-multiplying both sides of this equation by the two-component Pauli wavefunction Ψ for a spin-1/2 particle, we see that the equation for a free electron may be written as:

(39) $\left(E-\vec{\sigma}\bullet\vec{p}\right)\left(E+\vec{\sigma}\bullet\vec{p}\right)\Psi=m^2\Psi$,

where Ψ is a two-component wavefunction.

Applying the Schrödinger correspondence rules (4), equation (39) becomes:

(40) $\left(i\partial_t+i\vec{\sigma}\bullet\nabla\right)\left(i\partial_t-i\vec{\sigma}\bullet\nabla\right)\Psi=m^2\Psi$.

Bearing in mind our extended discussion of "Difficulty #1" in §4, we seek a wave equation which is of *first* order in the time derivative. Relativistic and aesthetic considerations suggest that the desired equation should be linear in ∇ also (cf. point (ii) made in the paragraph immediately after equation (9)). Please keep these comments in

---

[16] J. J. Sakurai, *Advanced Quantum Mechanics*, Addison-Wesley, Reading (1967), pp. 78-81.

[17] Bearing in mind the Schrödinger correspondence rule (4), $\vec{p}\times\vec{p}=i\nabla\times i\nabla=-\nabla\times\nabla$. Any elementary text on vector analysis will contain the identity $\left(\nabla\times\nabla\right)\Psi=0$, from which we see that $\vec{p}\times\vec{p}=-\nabla\times\nabla=0$.

[18] We have made use of the fact that E commutes with $\vec{\sigma}\bullet\vec{p}$, as can be seen if you expand out the left side of (38). If this point is unclear to you, please ask me to explain!

the back of your mind as we work through the derivation which forms the rest of this section.

Divide both sides of (40) by m:

$$(41)\quad \left(i\partial_t + i\vec{\sigma}\bullet\nabla\right)\left\{\frac{1}{m}\left(i\partial_t - i\vec{\sigma}\bullet\nabla\right)\Psi\right\} = m\Psi .$$

If we define the quantity in braces to be $\Psi^{(R)}$, we conclude that the second-order equation (40) is equivalent to a pair of first-order equations:

$$(42)\quad \begin{cases}\left(i\partial_t + i\vec{\sigma}\bullet\nabla\right)\Psi^{(R)} = m\Psi \\ \left(i\partial_t - i\vec{\sigma}\bullet\nabla\right)\Psi = m\Psi^{(R)}\end{cases}.$$

We have therefore achieved our aim of an equation which is first order in time, by doubling the number of wavefunctions we need to consider! Thus, rather than a single two-component Pauli wavefunction, we have to contend with a pair of two-component wavefunctions. This is equivalent to a single four-component wavefunction, about which more will be said later. For the moment, take the sum and the difference of each of equations (42), respectively, to give:

(42½)
$$\begin{cases}\left(i\partial_t + i\vec{\sigma}\bullet\nabla\right)\Psi^{(R)} + \left(i\partial_t - i\vec{\sigma}\bullet\nabla\right)\Psi = m\left(\Psi + \Psi^{(R)}\right) \\ \left(i\partial_t + i\vec{\sigma}\bullet\nabla\right)\Psi^{(R)} - \left(i\partial_t - i\vec{\sigma}\bullet\nabla\right)\Psi = m\left(\Psi - \Psi^{(R)}\right)\end{cases}$$

Re-arrange a little to give:

(43)
$$\begin{cases}i\partial_t\left(\Psi^{(R)} + \Psi\right) + i\vec{\sigma}\bullet\nabla\left(\Psi^{(R)} - \Psi\right) = m\left(\Psi + \Psi^{(R)}\right) \\ i\partial_t\left(\Psi^{(R)} - \Psi\right) + i\vec{\sigma}\bullet\nabla\left(\Psi^{(R)} + \Psi\right) = m\left(\Psi - \Psi^{(R)}\right)\end{cases}$$

Now let:

$$(44)\quad \Psi_A \equiv \Psi^{(R)} + \Psi, \quad \Psi_B \equiv \Psi^{(R)} - \Psi .$$

Therefore (43) becomes:

$$(45)\begin{cases}i\partial_t\Psi_A + i\vec{\sigma}\bullet\nabla\Psi_B = m\Psi_A \\ i\partial_t\Psi_B + i\vec{\sigma}\bullet\nabla\Psi_A = -m\Psi_B\end{cases}$$

This can be written in matrix form:

$$(46)\begin{bmatrix} i\partial_t & i\vec{\sigma}\bullet\nabla \\ i\vec{\sigma}\bullet\nabla & i\partial_t\end{bmatrix}\begin{bmatrix}\Psi_A \\ \Psi_B\end{bmatrix} = -m\begin{bmatrix}-\Psi_A \\ \Psi_B\end{bmatrix}.$$

Pull all terms over to the left-hand side:

$$\begin{bmatrix} i\partial_t & i\vec{\sigma}\bullet\nabla \\ i\vec{\sigma}\bullet\nabla & i\partial_t\end{bmatrix}\begin{bmatrix}\Psi_A \\ \Psi_B\end{bmatrix} + m\begin{bmatrix}-\Psi_A \\ \Psi_B\end{bmatrix} = 0$$

$$(47)\quad \begin{bmatrix} i\partial_t & i\vec{\sigma}\bullet\nabla \\ i\vec{\sigma}\bullet\nabla & i\partial_t\end{bmatrix}\begin{bmatrix}\Psi_A \\ \Psi_B\end{bmatrix} + m\begin{bmatrix}-1 & 0 \\ 0 & 1\end{bmatrix}\begin{bmatrix}\Psi_A \\ \Psi_B\end{bmatrix} = 0 .$$

$$\begin{bmatrix} i\partial_t & i\vec{\sigma}\bullet\nabla \\ i\vec{\sigma}\bullet\nabla & i\partial_t\end{bmatrix}\begin{bmatrix}\Psi_A \\ \Psi_B\end{bmatrix} - m\begin{bmatrix}1 & 0 \\ 0 & -1\end{bmatrix}\begin{bmatrix}\Psi_A \\ \Psi_B\end{bmatrix} = 0$$

Therefore:

$$(48)\quad \left\{\begin{bmatrix} i\partial_t & i\vec{\sigma}\bullet\nabla \\ i\vec{\sigma}\bullet\nabla & i\partial_t\end{bmatrix} - m\begin{bmatrix}1 & 0 \\ 0 & -1\end{bmatrix}\right\}\begin{bmatrix}\Psi_A \\ \Psi_B\end{bmatrix} = 0 .$$

Note that the first matrix in the braces of (48) can be written as:

(49)
$$\begin{bmatrix} i\partial_t & i\vec{\sigma}\bullet\nabla \\ i\vec{\sigma}\bullet\nabla & i\partial_t\end{bmatrix} = \begin{bmatrix} i\partial_t & 0 \\ 0 & i\partial_t\end{bmatrix} + \begin{bmatrix} 0 & i\vec{\sigma}\bullet\nabla \\ i\vec{\sigma}\bullet\nabla & 0\end{bmatrix}$$
$$= i\partial_t\begin{bmatrix}1 & 0 \\ 0 & 1\end{bmatrix} + i\begin{bmatrix}0 & \vec{\sigma} \\ \vec{\sigma} & 0\end{bmatrix}\bullet\nabla$$
$$= i\partial_t + i\begin{bmatrix}0 & \vec{\sigma} \\ \vec{\sigma} & 0\end{bmatrix}\bullet\nabla .$$

and so (48) becomes:

$$(50)\quad \left\{i\partial_t + i\begin{bmatrix}0 & \vec{\sigma} \\ \vec{\sigma} & 0\end{bmatrix}\bullet\nabla - m\begin{bmatrix}1 & 0 \\ 0 & -1\end{bmatrix}\right\}\begin{bmatrix}\Psi_A \\ \Psi_B\end{bmatrix} = 0 .$$

Now let:

$$(51)\quad \boldsymbol{\alpha} \equiv \begin{bmatrix}0 & \vec{\sigma} \\ \vec{\sigma} & 0\end{bmatrix}, \quad \beta \equiv \begin{bmatrix}1 & 0 \\ 0 & -1\end{bmatrix}, \quad \Psi \equiv \begin{bmatrix}\Psi_A \\ \Psi_B\end{bmatrix}.$$

(Note that the Ψ appearing in equation (51) is a four-component wavefunction, since both Ψ_A and Ψ_B are two-component Pauli wavefunctions.) Equation (51) allows us to re-write (50) as:

$$(52)\quad \left(i\partial_t + i\boldsymbol{\alpha}\bullet\nabla - \beta m\right)\Psi = 0 .$$

Bearing in mind the Schrödinger correspondence rules (4), we obtain our final result:

$$(53)\quad \boxed{\left(E - \boldsymbol{\alpha}\bullet\vec{p} - \beta m\right)\Psi = 0} .$$

Equation (53) is the famous *Dirac equation*, for the case of an electron in free space. Later, we will derive a more general form of the Dirac equation which is appropriate for an electron in an electromagnetic field.

The operators $\boldsymbol{\alpha} \equiv (\alpha_x, \alpha_y, \alpha_z)$ and β which appear in the Dirac equation (53) satisfy the so-called "Clifford algebra":

(54a) $\alpha_x^2 = \alpha_y^2 = \alpha_z^2 = 1$

(54b) $\beta^2 = 1$

(54c) $\alpha_k \alpha_m + \alpha_m \alpha_k = 0, \; k \neq m$

(54d) $\alpha_k \beta + \beta \alpha_k = 0$

where k and m take on the values x, y, z.

►*Exercise 6: Prove (54). Compare the results to those of exercise 4.*

►*Exercise 7: Show that* $\boldsymbol{\alpha}$ *and* β *are Hermitian,* i.e. that $\boldsymbol{\alpha}^\dagger = \boldsymbol{\alpha}$ and $\beta^\dagger = \beta$, *and then interpret this result. Hint: See footnote 22.*

We have twice mentioned the fact that the Dirac wavefunction has four components. We close this section by making these statements more concrete. We pointed out that wavefunction Ψ in (39) is a two-component Pauli wavefunction Ψ for a spin-1/2 particle, such as you would have met in your previous studies of spin (intrinsic) angular momentum. Explicitly, this two-component Pauli wavefunction is a function of position $\vec{r}$, time t and a spin variable σ:

$$\Psi(\vec{r}, t, \sigma) = \begin{bmatrix} \Psi_{1/2}(\vec{r}, t) \\ \Psi_{-1/2}(\vec{r}, t) \end{bmatrix} = \Psi_{1/2}(\vec{r}, t) \begin{bmatrix} 1 \\ 0 \end{bmatrix} + \Psi_{-1/2}(\vec{r}, t) \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

where the "spinors" $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$ are respectively equal to the "spin up" and "spin down" vectors of the spin-space, and $\Psi_{1/2}(\vec{r}, t)$ and $\Psi_{-1/2}(\vec{r}, t)$ may be thought of as the "spin up" and "spin down" components of the two-component Pauli wavefunction[19]. Bearing this in mind, we see that each of the two components of the Dirac wavefunction Ψ in (51) is itself a two-component wavefunction. Explicitly, the Dirac wavefunction can be written as:

(55) $$\Psi \equiv \begin{bmatrix} \Psi_A \\ \Psi_B \end{bmatrix} = \begin{bmatrix} \Psi_1 \\ \Psi_2 \\ \Psi_3 \\ \Psi_4 \end{bmatrix}$$

where Ψ_A and Ψ_B are two-component Pauli wavefunctions:

$$\Psi_A = \begin{bmatrix} \Psi_1 \\ \Psi_2 \end{bmatrix}, \quad \Psi_B = \begin{bmatrix} \Psi_3 \\ \Psi_4 \end{bmatrix},$$

and Ψ_1, Ψ_2, Ψ_3, Ψ_4 are single-component complex wavefunctions. The Dirac wavefunctions are known by various names, including "Dirac spinors", "bi-spinors" and "four-component wave functions"[20].

§5.3 Current and probability density[21]

To derive expressions for the current and probability density in the Dirac theory, we first introduce the Hermitian conjugate[22] wave functions, which are denoted by $\Psi^\dagger$ ("Psi dagger"):

(56) $$\Psi^\dagger \equiv \left[\Psi_1^*, \quad \Psi_2^*, \quad \Psi_3^*, \quad \Psi_4^*\right].$$

Pre-multiply both sides of (52) by $\Psi^\dagger$:

(57) $$\Psi^\dagger \left(i\partial_t + i\boldsymbol{\alpha} \bullet \nabla - \beta m\right)\Psi = 0.$$

Similarly, take the Hermitian conjugate of (52), i.e.:

(58) $$\Psi^\dagger \left(-i\partial_t - i\boldsymbol{\alpha}^\dagger \bullet \nabla - \beta^\dagger m\right) = 0$$

[19] See, for example, section 6.8 of B. H. Bransden and C. J. Jochain, *Introduction to Quantum Mechanics*, Longman Scientific and Technical, Essex (1989)..

[20] Note that the fact that this has four components has nothing to do with the four-dimensional nature of space and time. If you would like to explore this point further, please ask me!

[21] J. D. Bjorken and S. D. Drell, *Relativistic Quantum Mechanics*, McGraw-Hill Book Company, New York (1964), pp. 8-9

[22] The "Hermitian conjugate" is a generalization of the notion of "complex conjugate". The Dirac wavefunction Ψ can be considered as a *matrix* of one column and four rows (see (55)). The Hermitian conjugate of this matrix, denoted by $\Psi^\dagger$, is obtained by taking the complex conjugate of the transposed matrix, as has been done in going from (55) to (56). This definition of Hermitian conjugate is evidently applicable to any $m \times n$ matrix. A matrix is "Hermitian" if it is equal to its Hermitian conjugate. It is a fundamental postulate of quantum mechanics that any dynamical variable may be represented by a Hermitian operator.

and post-multiply by Ψ to give:

(58) $\Psi^\dagger\left(-i\partial_t - i\boldsymbol{\alpha}^\dagger \bullet \nabla - \beta^\dagger m\right)\Psi = 0$.

Now, we saw in exercise 7 that $\boldsymbol{\alpha}$ and β are Hermitian, i.e. that $\boldsymbol{\alpha}^\dagger = \boldsymbol{\alpha}$ and $\beta^\dagger = \beta$. Therefore (58) becomes:

(59) $\Psi^\dagger\left(-i\partial_t - i\boldsymbol{\alpha} \bullet \nabla - \beta m\right)\Psi = 0$.

►*Exercise 7½: In what way are equations (57) and (59) analogous to equations (18a) and (18b), which were used in deriving expressions for the probability density and probability current associated with the Klein-Gordon equation? Make use of this analogy, then, by adding (59) to (57) and hence arriving at the following expressions for probability density and current in the Dirac theory:*

(60a)

$$P = \Psi^\dagger\Psi \equiv \begin{bmatrix}\Psi_1^*, & \Psi_2^*, & \Psi_3^*, & \Psi_4^*\end{bmatrix}\begin{bmatrix}\Psi_1\\ \Psi_2\\ \Psi_3\\ \Psi_4\end{bmatrix} = \sum_{j=1}^{4}\left|\Psi_j\right|^2,$$

(60b) $\vec{j} = \Psi^\dagger\boldsymbol{\alpha}\Psi$.

You may find the following vector identity helpful:

(61) $\nabla \bullet \left(A\vec{B}\right) = \nabla A \bullet \vec{B} + A\nabla \bullet \vec{B}$

Interpret your results.

The probability density in (60a) is evidently positive definite, which overcomes one of the difficulties mentioned in our studies of the Klein-Gordon equation.

►*Exercise 8: Show that the so-called "gamma matrices":*

(62) $\gamma_0 \equiv \beta, \quad \gamma_1 \equiv \beta\alpha_x, \quad \gamma_2 \equiv \beta\alpha_y, \quad \gamma_3 \equiv \beta\alpha_z$

allow the algebra (54) to be written in the compact form:

(63) $\gamma_\mu\gamma_\nu + \gamma_\nu\gamma_\mu = 2g_{\mu\nu}, \quad \mu = 0,1,2,3, \quad \nu = 0,1,2,3,$

where $g_{\mu\nu}$ *is the flat space-time metric of special relativity:*

(64) $g_{\mu\nu} = \begin{bmatrix}1 & 0 & 0 & 0\\ 0 & -1 & 0 & 0\\ 0 & 0 & -1 & 0\\ 0 & 0 & 0 & -1\end{bmatrix}$.

Show also that the Hermitian property of the $\boldsymbol{\alpha}$ *and* β *matrices may be written as:*

(65) $\gamma_\mu^\dagger = \gamma_0\gamma_\mu\gamma_0$.

The gamma matrices are often used in studies of Quantum Field Theory which involve Quantum Electrodynamics (of which we will make a brief study in the next chapter) and the famous Feynman diagrams. These gamma matrices have a number of deep symmetry properties which are of both fundamental interest and practical utility in aspects of modern theoretical physics[23] .

§5.4 Dirac equation for an electron in the presence of an electromagnetic field[24]

We state without proof that, in the presence of an electromagnetic field described by a scalar potential φ and a vector potential $\vec{A}$, the Schrödinger correspondence rules (4) become:

(66) $E \to i\partial_t - e\varphi, \quad \vec{p} \to -i\nabla - e\vec{A}$,

where e is the electron charge. Evidently, our generalised correspondence rules reduce to the familiar correspondence rules (4) when the scalar and vector potentials of the electromagnetic field go to zero.

Now, the free-space Dirac equation (46) may be written in the form:

$$(67)\begin{bmatrix}E & -\vec{\sigma} \bullet \vec{p}\\ -\vec{\sigma} \bullet \vec{p} & E\end{bmatrix}\begin{bmatrix}\Psi_A\\ \Psi_B\end{bmatrix} = -m\begin{bmatrix}-\Psi_A\\ \Psi_B\end{bmatrix}.$$

To obtain the Dirac equation in the presence of an electromagnetic field, we make use of (66) to give:

$$(68)\begin{bmatrix}i\partial_t - e\varphi & \vec{\sigma} \bullet \left(i\nabla + e\vec{A}\right)\\ \vec{\sigma} \bullet \left(i\nabla + e\vec{A}\right) & i\partial_t - e\varphi\end{bmatrix}\begin{bmatrix}\Psi_A\\ \Psi_B\end{bmatrix} = m\begin{bmatrix}\Psi_A\\ -\Psi_B\end{bmatrix}.$$

[23] If interested, see sections A.2 and A.3 of F. Mandl and G. Shaw, *Quantum Field Theory (Revised Edition)*, John Wiley and Sons, Chichester (1993).

[24] J. J. Sakurai, *Advanced Quantum Mechanics*, Addison-Wesley, Reading (1967), p. 85.

This may be written as a pair of coupled equations:

$$(69)\quad \begin{cases} (i\partial_t - e\varphi - m)\Psi_A + \vec{\sigma}\bullet\left(i\nabla + e\vec{A}\right)\Psi_B = 0 \\ \vec{\sigma}\bullet\left(i\nabla + e\vec{A}\right)\Psi_A + (i\partial_t - e\varphi + m)\Psi_B = 0 \end{cases}.$$

Assume that both the vector potential $\vec{A}$ and the scalar potential φ are independent of time. Further, assume that both Ψ_A and Ψ_B are eigenfunctions of the energy operator with energy E_0, i.e.:

$$(70)\quad \begin{cases} \Psi_A(\vec{r},t) = \tilde{\Psi}_A(\vec{r})\exp(-iE_0 t), \\ \Psi_B(\vec{r},t) = \tilde{\Psi}_B(\vec{r})\exp(-iE_0 t). \end{cases}$$

►*Exercise 9: Verify that the wavefunctions in (70) are indeed eigenfunctions of the energy operator. Show that the probability density of the wavefunction in (70) is independent of time, and interpret this result.*

Using our assumption, equations (69) become:

$$(71)\quad \begin{cases} \vec{\sigma}\bullet\left(i\nabla + e\vec{A}\right)\Psi_B = (-E_0 + e\varphi + m)\Psi_A \\ \vec{\sigma}\bullet\left(i\nabla + e\vec{A}\right)\Psi_A = (-E_0 + e\varphi - m)\Psi_B \end{cases}.$$

Use the second of these equations to solve for Ψ_B:

$$(72)\quad \Psi_B = \frac{1}{-E_0 + e\varphi - m}\vec{\sigma}\bullet\left(i\nabla + e\vec{A}\right)\Psi_A,$$

thus yielding an expression which can be inserted into the first of equations (71) to give:

$$(73)\quad \vec{\sigma}\bullet\left(i\nabla + e\vec{A}\right)\frac{1}{-E_0 + e\varphi - m}\vec{\sigma}\bullet\left(i\nabla + e\vec{A}\right)\Psi_A = (-E_0 + e\varphi + m)\Psi_A.$$

►*Exercise 10: Evidently, equation (73) involves the only the "upper" component Ψ_A of the Dirac wavefunction (cf. (55)). It is reasonable to assume - or, at least, to take as a first working hypothesis - that the single partial differential equation (73) for Ψ_A is easier to solve than the pair of coupled differential equations (71) which "mix together" both Ψ_A and Ψ_B. In this context, it is natural to seek a single differential equation for Ψ_B, namely:*

$$(74)\quad \vec{\sigma}\bullet\left(i\nabla + e\vec{A}\right)\frac{1}{(-E_0 + e\varphi + m)}\vec{\sigma}\bullet\left(i\nabla + e\vec{A}\right)\Psi_B = (-E_0 + e\varphi - m)\Psi_B.$$

Please derive this result.

The good thing about equations (73) and (74) is that they are uncoupled, i.e. one is an equation for Ψ_A alone and the other is an equation for Ψ_B alone. This may be compared to (71), where we had a pair of equations, both of which contained a mixture of Ψ_A and Ψ_B. Uncoupled equations are typically much easier to deal with than coupled equations!

§5.5 Non-relativistic limit of Dirac equation for an electron in an electromagnetic field[25]

So far, we have developed a good deal of formalism regarding the Dirac equation. But what does it all mean? In the context of this question, we would like here to make some contact with our previous studies on non-relativistic quantum mechanics. Newtonian mechanics can be considered as a special limiting case of Einsteinian relativistic mechanics; similarly, the non-relativistic Schrödinger equation should be a special case of the relativistic Dirac equation. In exploring the non-relativistic limit of the Dirac equation, we shall have found a conceptual bridge which leads us from the less familiar (i.e. the Dirac equation) to the more familiar (i.e. the Schrödinger equation). This conceptual bridge will help us to deepen our understanding of the Dirac equation[26].

How do we quantify the notion of a "non-relativistic limit"? The answer to this question is provided by the relativistic relation (6), which states that the squared energy E^2 of a particle is given by the sum of its rest mass energy m^2 and its energy of motion $|\vec{p}|^2$. In the non-relativistic limit, the energy of motion is much smaller than the rest mass energy, allowing us to make the approximation:

$$(75a)\quad E_0 \approx m = \text{rest mass energy}\,(E_0 = mc^2, c = 1).$$

[25] J. J. Sakurai, *Advanced Quantum Mechanics*, Addison-Wesley, Reading (1967), pp. 85-86.

[26] Cf. the second half of §4.3, where we used the non-relativistic limit of the Klein-Gordon equation to assist us in its study.

Since we have complicated our lives by allowing electromagnetic fields to push the electrons around, the taking of the non-relativistic limit also requires us to assume that the electrostatic energy is much less than the rest-mass energy:

(75b) $|e\varphi| << m$.

We have therefore answered our question of what constitutes the non-relativistic limit with the answer, "In the non-relativistic limit, the rest mass energy dominates over other forms of energy such as the energy of motion and the electrostatic energy".

We can now make use of the non-relativistic approximations (75a) and (75b) to rewrite (73) as:

$$(76)\quad \vec{\sigma}\bullet\left(i\nabla+e\vec{A}\right)\frac{1}{-2m}\vec{\sigma}\bullet\left(i\nabla+e\vec{A}\right)\Psi_A \approx \left(e\varphi-\{E_0-m\}\right)\Psi_A .$$

The quantity in braces is the difference between the energy and the rest-mass energy, a useful and meaningful quantity which we denote by $\widetilde{E}$:

(77) $\widetilde{E}\equiv E_0-m$.

Thus (76) may be re-written as:

(78) $\vec{\sigma}\bullet\left(i\nabla+e\vec{A}\right)\vec{\sigma}\bullet\left(i\nabla+e\vec{A}\right)\Psi_A \approx 2m\left(\widetilde{E}-e\varphi\right)\Psi_A$.

We focus on the operator $\vec{\sigma}\bullet\left(i\nabla+e\vec{A}\right)\vec{\sigma}\bullet\left(i\nabla+e\vec{A}\right)$ which appears on the left side of this equation. From identity (31), we have:

$$(79)\quad \begin{aligned}&\vec{\sigma}\bullet\left(i\nabla+e\vec{A}\right)\vec{\sigma}\bullet\left(i\nabla+e\vec{A}\right)\\ &=\left(i\nabla+e\vec{A}\right)^2+i\vec{\sigma}\bullet\left(\left(i\nabla+e\vec{A}\right)\times\left(i\nabla+e\vec{A}\right)\right)\\ &=\left(i\nabla+e\vec{A}\right)^2+i\vec{\sigma}\bullet\left\{\begin{matrix}i\nabla\times i\nabla+i\nabla\times e\vec{A}\\ +e\vec{A}\times i\nabla+e\vec{A}\times e\vec{A}\end{matrix}\right\}\end{aligned}$$

The first and fourth terms inside the braces are both zero, and so (79) becomes:

$$(80)\quad \begin{aligned}&\vec{\sigma}\bullet\left(i\nabla+e\vec{A}\right)\vec{\sigma}\bullet\left(i\nabla+e\vec{A}\right)\\ &=\left(i\nabla+e\vec{A}\right)^2-e\vec{\sigma}\bullet\left\{\nabla\times\vec{A}+\vec{A}\times\nabla\right\}\end{aligned}$$

Next, we make use of the identity:

(81) $\nabla\times\vec{A}+\vec{A}\times\nabla=\left(\nabla\times\vec{A}\right)=\vec{B}$,

allowing us to reduce (80) to:

(82) $$\vec{\sigma}\bullet\left(i\nabla+e\vec{A}\right)\vec{\sigma}\bullet\left(i\nabla+e\vec{A}\right)=\left(i\nabla+e\vec{A}\right)^2-e\vec{\sigma}\bullet\vec{B}.$$

►*Exercise 11: Please prove equation (81)*[27].

We now substitute (82) into (78):

(83) $\left(\left(i\nabla+e\vec{A}\right)^2-e\vec{\sigma}\bullet\vec{B}\right)\Psi_A \approx 2m\left(\widetilde{E}-e\varphi\right)\Psi_A$,

and then re-arrange a little to end up with:

(84) $\left(\frac{1}{2m}\left(-i\nabla-e\vec{A}\right)^2-\frac{e}{2m}\vec{\sigma}\bullet\vec{B}+e\varphi\right)\Psi_A \approx \widetilde{E}\Psi_A$.

This is the (non-relativistic) Schrödinger-Pauli equation for an electron in the presence of an electromagnetic field, with the said electromagnetic field being described by the electric potential φ and the magnetic field $\vec{B}=\nabla\times\vec{A}$.

If we "switch off" the electromagnetic field:

(85) $\varphi\to 0,\quad \vec{A}\to 0\Rightarrow\vec{B}=\nabla\times\vec{A}\to 0$

then (84) becomes:

(86) $\frac{-\nabla^2}{2m}\Psi_A \approx \widetilde{E}\Psi_A$,

which is the familiar time-independent Schrödinger equation.

We have therefore achieved our aim of making contact between the non-relativistic limit of the Dirac equation, and the Schrödinger equation.

How does this help us in an interpretive sense? We have shown that, in the non-relativistic limit, the upper component Ψ_A of the Dirac wavefunction (cf. (55)) obeys the non-relativistic equations (84) or (86), depending on whether or not there is an electromagnetic field present. Thus the upper

[27] Hint: You might well ask, "What is the difference between $\nabla\times\vec{A}$ and $(\nabla\times\vec{A})$?" The answer lies in the fact that (81) is an operator equation, i.e. an equation which involves operators. For example, consider the apparently-bizarre statement $\nabla\times\vec{A}\neq(\nabla\times\vec{A})$. Remember that the quantities in this apparently-bizarre statement are operators; operators operate on something - say, some function Φ. Thus we can post-multiply both sides of our statement by Φ to give $\nabla\times\vec{A}\Phi\neq(\nabla\times\vec{A})\Phi$, i.e. $\nabla\times\left(\vec{A}\Phi\right)\neq(\nabla\times\vec{A})\Phi$.

component Ψ_A of the Dirac wavefunction converges to the non-relativistic Schrödinger-Pauli wavefunction in the non-relativistic limit.

It is natural to ask, "What about the lower component Ψ_B of the Dirac wavefunction? How does it behave in the non-relativistic limit?" We devote the remainder of this section to answering this question[28].

Take equation (72), then make approximation (75a) to give:

$$(87)\quad \Psi_B = \frac{1}{-2m+e\varphi}\vec{\sigma}\bullet\left(i\nabla+e\vec{A}\right)\Psi_A .$$

Now, approximation (75b) means that the first term in the denominator is much larger than the second term, and so this second term can be discarded to leave:

$$(88)\quad \Psi_B \approx \frac{1}{-2m}\vec{\sigma}\bullet\left(i\nabla+e\vec{A}\right)\Psi_A .$$

The following argument is a crude "back of the envelope" estimate, which I am happy to refine further if pressed by the aesthetes among you! Crudely, then, the operator on the right-side of (88) is the product of momentum and the $\vec{\sigma}$, divided by the rest mass. Since all elements of $\vec{\sigma}$ are of the order of 1 in size, the operator on the right-side of (88) is, crudely speaking, the momentum divided by the rest mass. In the non-relativistic limit, the momentum is much smaller than the rest mass (equation (6) implies that, in natural units, energy, mass and momentum all have the same units; in the non-relativistic limit, we have already argued that $|\vec{p}|^2 << m^2$, which implies that $p/m << 1$). Therefore:

$$(89)\quad \Psi_B << \Psi_A .$$

We conclude that, in the non-relativistic limit, Ψ_A converges to the non-relativistic Schrödinger-Pauli wavefunction in the non-relativistic limit, while Ψ_B becomes negligible in comparison with Ψ_A. In this context, note that Ψ_A and Ψ_B are sometimes respectively called the "large" and "small" components of the Dirac wavefunction.

[28] This discussion is based on J. J. Sakurai, *Advanced Quantum Mechanics*, Addison-Wesley, Reading (1967), p. 86, and A. Messiah, *Quantum Mechanics (volume II)*, North-Holland Publishing Company, Amsterdam (1961), p. 934.

§5.6 Approximate relativistically-corrected equation for an electrostatic problem[29,30]

For all of this chapter, we have worked in the so-called "natural units" which were introduced in §2. For the purposes of this section, however, we will explicitly include the factors of $\hbar$ and c in our equations. Our reason for doing this is that we want to calculate the smallest nonzero relativistic corrections as predicted by the Dirac theory. To do this, we need to keep track of the various powers of c in our equations.

Accordingly, we put the appropriate factors of $\hbar$ and c into (71):

$$(90)\quad \begin{cases} \vec{\sigma}\bullet\left(-i\hbar\nabla-\dfrac{e\vec{A}}{c}\right)\Psi_B = \dfrac{1}{c}\left(E_0-e\varphi-mc^2\right)\Psi_A \\ \vec{\sigma}\bullet\left(-i\hbar\nabla-\dfrac{e\vec{A}}{c}\right)\Psi_A = \dfrac{1}{c}\left(E_0-e\varphi+mc^2\right)\Psi_B \end{cases} .$$

In the previous section, we went from the exact equation (73) to the approximate equation (76) by making approximations (75a) and (75b). This amounted to writing the operator $1/(-E_0+e\varphi-m)$ in (73) as:

$$(91)\quad \frac{1}{-E_0+e\varphi-m} \approx \frac{1}{-m+e\varphi-m} \approx \frac{1}{-2m},$$

which leads directly from (73) to (76).

In this section, we show how a less crude approximation than (91) leads to us to the promised relativistic corrections for an electrostatic problem.

For the remainder of this section, we consider the electrostatic case where the vector potential $\vec{A}$ is zero. As we shall see, our calculations will be difficult enough without the complications of taking into account the vector potential!

Solve the second equation of (90) for Ψ_B and then substitute into the first equation, to give:

$$(92)\quad \begin{aligned} &\hbar^2(\vec{\sigma}\bullet\nabla)\frac{c^2}{E_0-e\varphi+mc^2}(\vec{\sigma}\bullet\nabla)\Psi_A \\ &\qquad = -\left(E_0-e\varphi-mc^2\right)\Psi_A \end{aligned} .$$

[29] This section is probably the hardest section in the course!

[30] J. J. Sakurai, *Advanced Quantum Mechanics*, Addison-Wesley, Reading (1967), pp. 86-88.

Now make use of (77), which in our non-natural units now reads $\widetilde{E} \equiv E_0 - mc^2$, to give:

(93)
$$\hbar^2(\vec{\sigma}\bullet\nabla)\frac{c^2}{2mc^2+\widetilde{E}-e\varphi}(\vec{\sigma}\bullet\nabla)\Psi_A = (e\varphi-\widetilde{E})\Psi_A .$$

Now make an approximate which is one order less crude than (91):

$$\frac{c^2}{2mc^2+\widetilde{E}-e\varphi} = \frac{c^2}{2mc^2}\frac{1}{1+\frac{\widetilde{E}-e\varphi}{2mc^2}}$$

(94)
$$= \frac{1}{2m}\left(1+\frac{\widetilde{E}-e\varphi}{2mc^2}\right)^{-1}$$
$$\approx \frac{1}{2m}\left(1-\frac{\widetilde{E}-e\varphi}{2mc^2}\right)$$

where we have made use of:

(95) $(1+\theta)^a \approx 1+a\theta, \quad |\theta| << 1$

Note that (94) reduces to (91) when $c \to \infty$. Using (94), equation (93) becomes:

(96)
$$\frac{\hbar^2}{2m}(\vec{\sigma}\bullet\nabla)\left(1-\frac{\widetilde{E}-e\varphi}{2mc^2}\right)(\vec{\sigma}\bullet\nabla)\Psi_A \approx (e\varphi-\widetilde{E})\Psi_A .$$

This can be written in the form:

(97a) $\widetilde{H}\Psi_A \approx \widetilde{E}\Psi_A$,

where:

(97b) $\widetilde{H} \equiv -\frac{\hbar^2}{2m}(\vec{\sigma}\bullet\nabla)\left(1-\frac{\widetilde{E}-e\varphi}{2mc^2}\right)(\vec{\sigma}\bullet\nabla)+e\varphi_A$.

Superficially, (97a) looks like a Schrödinger equation, but there are at least two difficulties with this interpretation.

Difficulty #1: The first problem is that (97b) contains $\widetilde{E}$, and so (97a) is not a true eigenvalue equation.

Difficulty #2: The second problem is that the wavefunction Ψ_A is not normalised, as the expression for probability density which we derived in (60a) implies that:

(98)
$$\iiint \Psi^\dagger \Psi \,dx\,dy\,dz = 1$$
$$\Rightarrow \iiint \left(\Psi_A^\dagger\Psi_A + \Psi_B^\dagger\Psi_B\right)dx\,dy\,dz = 1 .$$

We now show how to overcome these difficulties.

►*Exercise 12: Please explain why Difficulty #2 was not considered in the discussions of §5.5.*

Take the second of equations (90), set $\vec{A}$ equal to zero, and then solve for Ψ_B, to give:

(99) $$\frac{-ic\hbar}{E_0 - e\varphi + mc^2}(\vec{\sigma}\bullet\nabla)\Psi_A = \Psi_B$$

►*Exercise 13: Prove equation (99), and then explain why this expression is useful.*

Since $\widetilde{E} \equiv E_0 - mc^2$,

(100) $$\frac{-ic\hbar}{\widetilde{E}-e\varphi+2mc^2}(\vec{\sigma}\bullet\nabla)\Psi_A = \Psi_B .$$

Again, the rest mass energy dominates and so the first two terms in the denominator can be scratched out:

(101a) $$\frac{1}{2mc}(\vec{\sigma}\bullet\vec{p})\Psi_A = \Psi_B ,$$

where use has been made of (4). The Hermitian conjugate[31] of this equation is:

(101b) $$\Psi_A^\dagger\frac{1}{2mc}(\vec{\sigma}\bullet\vec{p}) = \Psi_B^\dagger ,$$

where we have made use of the fact that $(\vec{\sigma}\bullet\vec{p})$ is Hermitian, i.e. $(\vec{\sigma}\bullet\vec{p})^\dagger = (\vec{\sigma}\bullet\vec{p})$. Multiplying together (101a) and (101b), we see that:

(102) $$\Psi_B^\dagger\Psi_B = \frac{1}{4m^2c^2}\Psi_A^\dagger(\vec{\sigma}\bullet\vec{p})(\vec{\sigma}\bullet\vec{p})\Psi_A .$$

[31] An earlier footnote discussed how to take the Hermitian conjugate of matrices. Here, we are faced with the problem of taking the Hermitian conjugate of products of operators. This is done by putting a dagger on all operators, reversing the order of all of these operators, and then replacing all complex quantities with their complex conjugates. For example, the Hermitian conjugate of iAB is $-iB^\dagger A^\dagger$.

It is left as an exercise to show that:

(103) $(\vec{\sigma}\bullet\vec{p})(\vec{\sigma}\bullet\vec{p}) = \vec{p}^{\,2}$,

where $\vec{p}^{\,2} \equiv \vec{p}\bullet\vec{p}$.

►*Exercise 14: Prove equation (103). You may find identity (31) helpful in this respect.*

Using (103), (102) becomes:

(104) $\Psi_B^{\dagger}\Psi_B = \dfrac{1}{4m^2c^2}\Psi_A^{\dagger}\vec{p}^{\,2}\Psi_A$.

Thus the normalisation condition (98) becomes:

(105) $\iiint \Psi_A^{\dagger}\left(1+\dfrac{\vec{p}^{\,2}}{4m^2c^2}\right)\Psi_A\,dx\,dy\,dz = 1$.

Now, to order $1/c^2$, the following is true:

(106) $\left(1+\dfrac{\vec{p}^{\,2}}{4m^2c^2}\right) \approx \left(1+\dfrac{\vec{p}^{\,2}}{8m^2c^2}\right)^2$.

(By "to order $1/c^2$", we mean: "The quantity c^{-1} is small, c^{-2} is smaller still, c^{-3} is even smaller *etc*. All quantities containing c^{-3}, c^{-4} *etc*. will be ignored as negligible for the purposes of the present calculation.") Therefore (105) can be written as:

(107) $\iiint \left(\Omega\Psi_A\right)^{\dagger}\left(\Omega\Psi_A\right)dx\,dy\,dz = 1$

where:

(108) $\Omega \equiv \left(1+\dfrac{\vec{p}^{\,2}}{8m^2c^2}\right)$.

To order $1/c^2$,

(109) $\Omega^{-1} \approx \left(1-\dfrac{\vec{p}^{\,2}}{8m^2c^2}\right)$.

Bearing (107) in mind, introduce the wavefunction:

(110) $\breve{\Psi} \equiv \Omega\Psi_A$

Thus (97a) becomes:

(111) $\tilde{H}\Omega^{-1}\breve{\Psi} \approx \tilde{E}\Omega^{-1}\breve{\Psi}$.

Pre-multiply both sides of this equation by Ω^{-1}:

(112) $\Omega^{-1}\tilde{H}\Omega^{-1}\breve{\Psi} \approx \tilde{E}\Omega^{-2}\breve{\Psi}$.

►*Exercise 15: Substitute (97b), (108) and (109) into equation (112). Expand everything out. The result will be a mess, but only keep terms to order $1/c^2$. Show that this gives (113) below, where the anti-commutator is defined by $[A,B]_+ \equiv AB+BA$.*

(113) $\left(\dfrac{\vec{p}^{\,2}}{2m}+e\varphi-\dfrac{\vec{p}^{\,4}}{8m^3c^2}-\left[\dfrac{\vec{p}^{\,2}}{8m^2c^2},e\varphi\right]_+-\dfrac{1}{4m^2c^2}(\vec{\sigma}\bullet\vec{p})(\tilde{E}-e\varphi)(\vec{\sigma}\bullet\vec{p})\right)\breve{\Psi} = \tilde{E}\left(1-\dfrac{\vec{p}^{\,2}}{4m^2c^2}\right)\breve{\Psi}$

On the right-hand side of this expression, we make use of the trivial anti-commutator $\tilde{E}\vec{p}^{\,2} = \frac{1}{2}[\tilde{E},\vec{p}^{\,2}]_+$ to give:

(114) $\left(\dfrac{\vec{p}^{\,2}}{2m}+e\varphi-\dfrac{\vec{p}^{\,4}}{8m^3c^2}-\left[\dfrac{\vec{p}^{\,2}}{8m^2c^2},e\varphi\right]_+-\dfrac{1}{4m^2c^2}(\vec{\sigma}\bullet\vec{p})(\tilde{E}-e\varphi)(\vec{\sigma}\bullet\vec{p})+\dfrac{[\tilde{E},\vec{p}^{\,2}]_+}{8m^2c^2}\right)\breve{\Psi} = \tilde{E}\breve{\Psi}$.

Next, merge the two anti-commutators in the left side of (114):

(115) $\left(\dfrac{\vec{p}^{\,2}}{2m}+e\varphi-\dfrac{\vec{p}^{\,4}}{8m^3c^2}-\dfrac{1}{8m^2c^2}\left[\vec{p}^{\,2},e\varphi-\tilde{E}\right]_+-\dfrac{1}{4m^2c^2}(\vec{\sigma}\bullet\vec{p})(\tilde{E}-e\varphi)(\vec{\sigma}\bullet\vec{p})\right)\breve{\Psi} = \tilde{E}\breve{\Psi}$.

We now pull out a common factor from the last two terms appearing in the big brackets above:

(116) $\left(\dfrac{\vec{p}^{\,2}}{2m}+e\varphi-\dfrac{\vec{p}^{\,4}}{8m^3c^2}+\dfrac{1}{8m^2c^2}\left\{\left[\vec{p}^{\,2},\tilde{E}-e\varphi\right]_+-2(\vec{\sigma}\bullet\vec{p})(\tilde{E}-e\varphi)(\vec{\sigma}\bullet\vec{p})\right\}\right)\breve{\Psi} = \tilde{E}\breve{\Psi}$.

Regarding the term $\vec{p}^{\,2}$ appearing in the anti-commutator, make use of (103) to give:

(116) $\left(\dfrac{\vec{p}^{\,2}}{2m}+e\varphi-\dfrac{\vec{p}^{\,4}}{8m^3c^2}+\dfrac{1}{8m^2c^2}\left\{\left[(\vec{\sigma}\bullet\vec{p})(\vec{\sigma}\bullet\vec{p}),\tilde{E}-e\varphi\right]_+-2(\vec{\sigma}\bullet\vec{p})(\tilde{E}-e\varphi)(\vec{\sigma}\bullet\vec{p})\right\}\right)\breve{\Psi} = \tilde{E}\breve{\Psi}$.

Evidently, the term in braces is of the form:

(117) $[A^2,B]_+ - 2ABA$,

where $A = (\vec{\sigma}\bullet\vec{p})$ and $B = \widetilde{E} - e\varphi$. In this context, we note the following useful identity:

(118) $[A^2,B]_+ - 2ABA = [A,[A,B]]$.

►*Exercise 16: Prove (118).*

This identity allows us to rewrite (116) as:

$$(119)\quad \left(\frac{\vec{p}^2}{2m} + e\varphi - \frac{\vec{p}^4}{8m^3c^2} + \frac{\left[(\vec{\sigma}\bullet\vec{p}),\left[(\vec{\sigma}\bullet\vec{p}),\widetilde{E}-e\varphi\right]\right]}{8m^2c^2}\right)\breve{\Psi} = \widetilde{E}\breve{\Psi}\,.$$

(As a check, note that this equation reduces to what it should when you set $c\to\infty$ and turn off the electric field by setting $\varphi\to 0$.)

Now we concentrate on the "internal commutator" in (119):

$$\begin{aligned}
&\left[(\vec{\sigma}\bullet\vec{p}),\widetilde{E}-e\varphi\right]\breve{\Psi}\\
&= \left[(\vec{\sigma}\bullet\vec{p}),-e\varphi\right]\breve{\Psi} \text{ ... } \widetilde{E} \text{ drops out}\\
&= -e\left[(\vec{\sigma}\bullet\vec{p}),\varphi\right]\breve{\Psi} \text{ ... factor out the - e}\\
&= -e\left((\vec{\sigma}\bullet\vec{p})\varphi - \varphi(\vec{\sigma}\bullet\vec{p})\right)\breve{\Psi} \text{ ...expand commutator}\\
(120)\quad &= ie\hbar\left((\vec{\sigma}\bullet\nabla)\varphi - \varphi(\vec{\sigma}\bullet\nabla)\right)\breve{\Psi} \text{ ... since } \vec{p} = -i\hbar\nabla\\
&= ie\hbar\left((\vec{\sigma}\bullet\nabla)\varphi\breve{\Psi} - \varphi(\vec{\sigma}\bullet\nabla)\breve{\Psi}\right) \text{... pull } \breve{\Psi} \text{ inside}\\
&= ie\hbar\left(\varphi(\vec{\sigma}\bullet\nabla)\breve{\Psi} + ((\vec{\sigma}\bullet\nabla)\varphi)\breve{\Psi} - \varphi(\vec{\sigma}\bullet\nabla)\breve{\Psi}\right)\\
&\quad \text{... we used the product rule to obtain the line above}\\
&= ie\hbar(\vec{\sigma}\bullet\nabla\varphi)\breve{\Psi} \text{ ... cancel first and third terms}\\
&= -ie\hbar(\vec{\sigma}\bullet\vec{E})\breve{\Psi} \text{ ... because } \vec{E} = -\nabla\varphi
\end{aligned}$$

Therefore,

(121) $\left[(\vec{\sigma}\bullet\vec{p}),\widetilde{E}-e\varphi\right] = -ie\hbar(\vec{\sigma}\bullet\vec{E})$.

Building on this result, it can be shown that:

$$(122)\quad \begin{aligned}&\left[(\vec{\sigma}\bullet\vec{p}),\left[(\vec{\sigma}\bullet\vec{p}),\widetilde{E}-e\varphi\right]\right]\\ &\quad = -e\hbar^2(\nabla\bullet\vec{E}) - 2e\hbar\vec{\sigma}\bullet(\vec{E}\times\vec{p})\end{aligned}.$$

►*Exercise 17: Prove (122). To help you with this calculation, you might find it useful to study some of the "tricks" used in equation (120).*

Using (122), (119) becomes:

$$(123)\quad \left(\frac{\vec{p}^2}{2m} + e\varphi - \frac{\vec{p}^4}{8m^3c^2} - \frac{e\hbar}{4m^2c^2}\vec{\sigma}\bullet(\vec{E}\times\vec{p}) - \frac{e\hbar^2}{8m^2c^2}(\nabla\bullet\vec{E})\right)\breve{\Psi} = \widetilde{E}\breve{\Psi}\,.$$

Equation (123) is the central result of this section. It is an approximate equation for relativistic electrostatics; it is approximate because we have only considered relativistic corrections up to order c^{-2}.

(a) When the speed of light is taken to infinity, equation (123) evidently reduces to the Schrödinger-Pauli equation (84), for the special case of zero vector potential $\vec{A}$.

(b) Bearing this in mind, the meaning of the first two terms in equation (123) is clear, with all remaining terms being relativistic corrections which disappear when the speed of light is made infinite.

(c) The third term in (123) is a relativistic correction to the kinetic energy, as can be seen from the expansion:

$$(124)\quad \sqrt{(mc^2)^2 + \vec{p}^2c^2} - mc^2 = \frac{\vec{p}^2}{2m} - \frac{\vec{p}^4}{8m^3c^2} + \ldots$$

►*Exercise 18: Prove (124).*

(d) The fourth term of (123) represents the spin interaction of the moving electron with the electric field, and is known as the "Thomas term" after L.H. Thomas who managed to derive this term two years *before* Dirac invented his famous equation! For the case of a central potential:

$$(125)\quad \vec{E} = -\nabla\varphi(r) = -\hat{r}\partial_r\varphi(r) = -\frac{\vec{r}}{r}\partial_r\varphi(r)$$

and so the Thomas term becomes:

$$(126)\quad \frac{e\hbar\partial_r\varphi(r)}{4m^2c^2r}\vec{\sigma}\bullet(\vec{r}\times\vec{p}).$$

Since the orbital angular momentum is given by:

(127) $\vec{L} = \vec{r}\times\vec{p}$,

the Thomas term is:

(128) $\dfrac{e\hbar\partial_r\varphi(r)}{4m^2c^2r}\vec{\sigma}\bullet\vec{L}$.

This provides a rigorous means of introducing the well-known spin-orbit force which you will have encountered in a previous course on quantum mechanics. This spin-orbit force is responsible for the fine structure of atomic energy levels.

(e) The last term in (123) is known as the Darwin term; we shall not consider it further, apart from commenting that one of the Maxwell equations implies that this term is proportional to the charge density.

(f) Using the third, fourth and fifth terms of (123) as the perturbation Hamiltonian, and the wavefunctions for the hydrogen atom in non-relativistic quantum mechanics as the unperturbed wavefunctions, we can compute the lowest-order relativistic corrections to the energy levels of the hydrogen atom. Since this calculation is based on first-order time-independent perturbation theory, we present only the result:

(129) $\Delta E = -\left(\dfrac{e^2}{4\pi\hbar c}\right)^2\left(\dfrac{e^2}{8\pi a_0}\right)\dfrac{1}{n^3}\left(\dfrac{1}{j+\frac{1}{2}}-\dfrac{3}{4n}\right)$,

which is added to the unperturbed energy levels:

(130) $E^{(0)} = \dfrac{e^2}{8\pi a_0 n^2}$.

APPENDIX A: NON-LOCALITY OF THE RELATIVISTIC SCHRÖDINGER EQUATION[32] (not examinable)

In this appendix, we demonstrate the non-local nature of the non-relativistic Schrödinger equation given by (9) in the main text.

Write the time-dependent wavefunction as a Fourier integral:

(A1) $\Psi(\vec{r},t) = (2\pi)^{-3/2}\iiint d^3\vec{p}\,\Psi(\vec{p},t)e^{i\vec{p}\bullet\vec{r}}$,

which may be read as saying that "$\Psi(\vec{r},t)$ is the inverse Fourier transform of $\Psi(\vec{p},t)$". Physically, equation (A1) expresses the wavefunction $\Psi(\vec{r},t)$ as a continuous sum (i.e. an integral) of plane waves / Fourier components $\exp(i\vec{p}\bullet\vec{r})$; in this sum, each of the infinitely many components $\exp(i\vec{p}\bullet\vec{r})$ are weighted by the infinitesimally-small factor $\Psi(\vec{p},t)d^3\vec{p}$.

We shall also have need for the other equation of what is known as a "Fourier transform pair", namely:

(A2) $\Psi(\vec{p},t) = (2\pi)^{-3/2}\iiint d^3\vec{r}\,\Psi(\vec{r},t)e^{-i\vec{p}\bullet\vec{r}}$.

This equation may be read as saying that "$\Psi(\vec{p},t)$ is the Fourier transform of $\Psi(\vec{r},t)$".

Substitute (A1) into the form (9) of the non-relativistic Schrödinger equation as given in the main text, to give:

(A3) $i\partial_t\Psi(\vec{r},t) = \dfrac{\sqrt{-\nabla^2+m^2}}{(2\pi)^{3/2}}\iiint d^3\vec{p}\,\Psi(\vec{p},t)e^{i\vec{p}\bullet\vec{r}}$.

Since $\sqrt{-\nabla^2+m^2}$ acts only on functions of the position vector $\vec{r}$, we may bring this operator inside the triple integral and allow it to "move through" all functions which do not contain $\vec{r}$. Thus we rewrite (A3) as:

(A4) $i\partial_t\Psi(\vec{r},t) = \dfrac{1}{(2\pi)^{3/2}}\iiint d^3\vec{p}\,\Psi(\vec{p},t)\sqrt{-\nabla^2+m^2}\,e^{i\vec{p}\bullet\vec{r}}$.

Next, we make use of the formula:

(A5) $\nabla^2 e^{i\vec{p}\bullet\vec{r}} = (i\vec{p}\bullet i\vec{p})e^{i\vec{p}\bullet\vec{r}} = -|\vec{p}|^2 e^{i\vec{p}\bullet\vec{r}}$

to rewrite (A4) as:

(A6) $i\partial_t\Psi(\vec{r},t) = \dfrac{1}{(2\pi)^{3/2}}\iiint d^3\vec{p}\,\Psi(\vec{p},t)\sqrt{|\vec{p}|^2+m^2}\,e^{i\vec{p}\bullet\vec{r}}$.

Substitute equation (A2) into (A6) to give:

(A7) $i\partial_t\Psi(\vec{r},t) = \dfrac{1}{(2\pi)^3}\times\iiint d^3\vec{p}\left\{\iiint d^3\vec{r}\,'\Psi(\vec{r}\,',t)e^{-i\vec{p}\bullet\vec{r}\,'}\right\}\sqrt{|\vec{p}|^2+m^2}\,e^{i\vec{p}\bullet\vec{r}}$

Note that the position vector in braces has been primed, since it is a dummy variable which is "integrated out" in the integral appearing within the braces; the position vector appearing *outside* the braces is not primed.

[32] R. H. Landau, *Quantum Mechanics II: A Second Course in Quantum Theory*, John Wiley & Sons, New York (1996), section 13.2.

Rearrange the terms appearing in (A7):

(A8)
$$i\partial_t \Psi(\vec{r},t) = \iiint d^3\vec{r}' \left\{ \frac{1}{(2\pi)^3} \iiint d^3\vec{p}\, e^{i\vec{p}\bullet(\vec{r}-\vec{r}')} \sqrt{|\vec{p}|^2 + m^2} \right\} \Psi(\vec{r}',t) .$$

Now let:

(A9) $$K(\vec{r},\vec{r}') \equiv \frac{1}{(2\pi)^3} \iiint d^3\vec{p}\, e^{i\vec{p}\bullet(\vec{r}-\vec{r}')} \sqrt{|\vec{p}|^2 + m^2} ,$$

allowing us to rewrite (A8) as:

(A10) $$i\partial_t \Psi(\vec{r},t) = \iiint K(\vec{r},\vec{r}') \Psi(\vec{r}',t) d^3\vec{r}' .$$

When written in this way, the non-locality of the relativistic Schrödinger equation becomes manifest.

To clarify this point, recall our earlier comment in footnote 9 that "non-local" means that the time-evolution of the wavefunction at a given point in space and at some given instant of time, is governed by the value of the wavefunction through all space, at that instant of time. This is precisely what equation (A10) says: the time evolution of the wavefunction (i.e. $\partial_t \Psi(\vec{r},t)$) at a given point $(\vec{r},t)$ in space and time depends on the wavefunction $\Psi(\vec{r}',t)$ at every point $\vec{r}'$ in space at that same instant of time t, via the integral $\iiint K(\vec{r},\vec{r}') \Psi(\vec{r}',t) d^3\vec{r}'$.

►*Exercise 19: Generalise the free-particle Klein-Gordon equation to allow for the presence of an electromagnetic field. Obtain the suitably-generalised equations for the probability density and probability current.*

► *Exercise 20: Find the free-particle solutions of the Klein-Gordon equation corresponding to well-defined momentum values (these are the "plane wave solutions" to the Klein-Gordon equation).*

► *Exercise 21: We learned in lectures that, since the free-particle Klein-Gordon equation is second order in time, the wavefunction at a given instant of time does not determine the wavefunction for all time. This suggests that we could work with a two-component wavefunction, thereby obtaining an equation first order in time, which in turn implies that the value of the two-component wavefunction at a given time determines the value of the two-component wavefunction for all time. Bearing these comments in mind, introduce an appropriate two-level wavefunction Ψ which allows you to recast the Klein-Gordon equation into the form of a Schrödinger equation: $i\partial_t \Psi = H\Psi$, where t denotes time and H is the Hamiltonian. Write down an explicit expression for this Hamiltonian, and then briefly interpret your result.*

►*Exercise 22: Find the energy levels of a π-mesic atom, that is, a system of a nucleus of charge Ze (Z = atomic number and e = absolute value of electron charge), which one may assume to be infinitely heavy, and a π^- meson. In this context, you will find it useful to review the non-relativistic calculation of the energy levels of the hydrogen atom using the Schrödinger equation, and then adapt this knowledge to the Klein-Gordon equation.*

►*Exercise 23: Find the solutions of the Dirac equation of a free particle corresponding to states of well-defined momentum (these are the "plane wave solutions" to the Dirac equation).*

► *Exercise 24: In the absence of a field, show that any solution of the Dirac equation is a solution of the Klein-Gordon equation, but that the converse is not true.*

CHAPTER 2
QUANTUM COMPUTERS

§7. Motivation

There are a number of reasons for inclusion of this topic. (a) I wished to include some truly contemporary material in these lectures. Physicists are often justly criticized for teaching "modern physics" courses whose latest results are decades old. In contrast, some of the results covered in this last section of the course will be less than a few years old. (b) The field of quantum computation gives a somewhat unexpected and beautiful application of some of the most fundamental concepts of quantum mechanics, such as superposition, entanglement, unitary evolution, coherent states, decoherence and so forth. It also gives some appreciation of the vast dimensionality

of the Hilbert spaces in which quantum systems live, spaces exponentially larger in dimension than their classical counterparts. Such are the concepts which underlie the massive parallelism of which quantum computers are capable. (c) In view of the influence which computing devices have over our everyday lives, it is important to outline what will potentially become a dominant technology in the mid-twenty-first century.

§8. Moore's law and the impending crisis of computation[33]

Starting with the invention of the transistor about fifty years ago, modern civilization has witnessed an approximately exponential decrease with time in the size of transistors used to construct computing devices. This exponential trend in miniaturization is known as Moore's law. If we extrapolate this trend into the future it becomes evident that metal-oxide field-effect transistors (MOSFETS) - the basic unit of modern computers - will run against fundamental limitations around the year 2015. This "crisis" is a simple consequence of fundamental quantum mechanics: when devices reach dimensions below about 30 nm, they will be comparable in size to the de-Broglie wavelength of the charge carriers and therefore quantum-mechanical effects will dominate. It is therefore evident that a new and explicitly quantum-mechanical paradigm of computation is needed to describe such computing devices.

§9. Classical and quantum computing[34]

In this section, we clarify a piece of confusion which may have arisen as you read the final two sentences of the previous section. You may rightly object to their implicit notion that contemporary computing machines are classical. This is because the theory of transistors contains a good deal of quantum mechanics, as those who have taken an elementary course in Electronics or Solid-State physics will know[35].

The answer to this objection is as follows. Information processing has hitherto been considered largely in classical terms, and modern digital computers are no exception to this rule. It is of no relevance to this argument that quantum mechanics has played a supporting role in the design of equipment to process such classical information.

A slightly forced analogy might help - despite the fact that you might need to invoke quantum mechanics to explain why water freezes when and how it does, this does not mean that an abacus made of ice cubes is a quantum computer.

§10. Classical and quantum information[36]

§10.1 Classical information

The basic unit of classical information processing is the *bit*. A classical bit (such as a memory element or a wire carrying a binary signal) is generally a macroscopic system such as a circuit on a breadboard or a digital computer's integrated circuit, and is described by one or more continuous parameters such as voltages. Within this parameter space two well-separated regions are chosen by the designer to represent 0 and 1, and signals are periodically restored toward these standard regions to prevent them from drifting away due to environmental influences, cross-talk, and finite manufacturing tolerances. A classical *n*-bit memory can exist in any of 2^n logical states, labelled 000...0 through to 111...1. Besides storing binary data, classical computers manipulate it; a sequence of Boolean operators (for example, AND and NOT) acting on the bits two at a time is sufficient to realize any deterministic transformation.

[33] D. C. Giatti, *Single electrons in silicon drops*, Nature **393** 516-517 (1998).
[34] C. H. Bennett and D. P. Vincenzo, *Quantum information and computation*, Nature **404** 247-255 (2000).
[35] If you are interested in pursuing this point see, for example, Chapters 5 and 6 ("Quantum theory of conduction" and "Semiconductor devices") of M. N. Rudden and J. Wilson, *Elements of solid-state physics*, John Wiley and Sons, Chichester (1993).
[36] C. H. Bennett and D. P. Vincenzo (2000).

§10.2 Quantum information

A quantum bit, or *qubit*, is typically embodied by a microscopic system such as an atom, a nuclear spin or a photon. The Boolean states 0 and 1 are represented by a fixed pair of reliably distinguishable states of the qubit (for example, horizontal and vertical photon polarizations: $|0\rangle = \leftrightarrow$, $|1\rangle = \updownarrow$). A qubit can also exist in a continuum of intermediate states or superpositions, represented mathematically as complex linear combinations of the basis states $|0\rangle$ and $|1\rangle$. For photons, these states correspond to other polarizations, for example $\nearrow\!\!\!\!\!\!\searrow = \left(|0\rangle + |1\rangle\right)/\sqrt{2}$, $\nwarrow\!\!\!\!\!\!\searrow = \left(|0\rangle - |1\rangle\right)/\sqrt{2}$ and $\circlearrowright = \left(|0\rangle + i|1\rangle\right)/\sqrt{2}$ (right circular polarization). Unlike the intermediate states of a classical bit (such as voltages between the standard 0 and 1 values) , these intermediate states cannot be reliably distinguished, even in principle, from the basis states. With regard to any measurement, the superposition $\alpha|0\rangle + \beta|1\rangle$ behaves like $|0\rangle$ with probability $|\alpha|^2$ and like $|1\rangle$ with probability $|\beta|^2$. More generally two quantum states are reliably distinguishable if and only if their vector representations are orthogonal : thus $\leftrightarrow$ and $\updownarrow$ are reliably distinguishable by one type of measurement and $\nearrow\!\!\!\!\!\!\searrow$ and $\nwarrow\!\!\!\!\!\!\searrow$ by another, but no measurement can reliably distinguish $\leftrightarrow$ and $\nearrow\!\!\!\!\!\!\searrow$.

A pair of qubits (for example two photons in different locations) is capable of existing in four Boolean states $|00\rangle, |01\rangle, |10\rangle, |11\rangle$, as well as all possible complex superpositions :

$$(173)\quad \frac{\alpha|00\rangle + \beta|01\rangle + \gamma|10\rangle + \delta|11\rangle}{\sqrt{|\alpha|^2 + |\beta|^2 + |\gamma|^2 + |\delta|^2}}$$

of them. These include states such as :

$$(174)\quad \frac{|00\rangle + |01\rangle}{\sqrt{2}} = |0\rangle \otimes \frac{|0\rangle + |1\rangle}{\sqrt{2}} = \leftrightarrow\!\!\!\!\!\!\nearrow\!\!\!\!\!\!\searrow$$

which is describable as a tensor product of states of the individual photons, as well as *maximally entangled* states such as :

$$(175)\quad \frac{|00\rangle + |11\rangle}{\sqrt{2}}$$

which admit no such description. Such entangled states correspond to a situation in which neither photon by itself has a definite state, even though the pair together does[37].

More generally, where a string of *n* classical bits could exist in any one of 2^n Boolean states 000...0 through 111...1, a string of *n* qubits can exist in any state of the form :

$$(176)\quad \Psi = \sum_{x=000\ldots0}^{x=111\ldots1} c_n |x\rangle,$$

where the c_x are complex numbers such that $\Sigma_x |c_x|^2 = 1$. Stated differently, a quantum state of *n* qubits is represented by a complex vector Ψ of unit length in a (Hilbert) space of 2^n dimensions, one for each classical state. *The exponentially large dimension of this space distinguishes quantum computers from classical computers, whose state is described by a number of parameters which grows only linearly with the size of the system. This is because classical systems, whether digital or analogue, can be completely described by*

[37] The most famous entangled state is probably Schrödinger's cat. Suppose a cat is concealed inside a box with a radioactive atom that may or may not decay during a certain time interval. If the decay of the atom is detected inside the box, a mechanism frees some prussic acid, which immediately kills the cat. The atom is a quantum object, so quantum mechanics applies and we can calculate the corresponding quantum state in the box. Because there is no contact between the cat and the outside world, then, according to quantum theory, the state inside the box must be described by a superposition of the cat being alive (atom not decayed) and the cat being dead (atom decayed) : Ψ = |atom decayed>|cat dead> + |atom not decayed>|cat alive>. Only opening the box - that is, making a measurement in the quantum-mechanical sense of the word - would reveal whether the cat is alive or dead. We see that the fate of the cat is inextricably interwoven with that of the atom. Because we do not know the state of the atom, we do not know of the well-being or otherwise of the cat. But if we know the cat is dead, then we can be certain the atom has decayed. This one-to-one correlation arises due to the *entanglement* of the states of the atom and cat in the wavefunction Ψ. Mathematically, the fact that the states "know" of each other shows up in the wavefunction describing the entangled state, which cannot be expressed as a product of the wavefunctions of the constituent states. This explicitly demonstrates the nonlocal character of quantum theory. (Taken from Rainer Blatt, *Push-button entanglement*, Nature **404** 231-232 (2000).)

separately describing the state of each part. The vast majority of quantum states, by contrast, are entangled, and admit no such description. The ability to preserve and manipulate entangled states is the distinguishing feature of quantum computing machines, responsible both for their power and for the difficulty of building them.

§11. The quantum circuit model[38]

Before introducing the quantum circuit model, we need to take a brief detour and describe the classical circuit model. Recall that a classical circuit can always be written solely with the three "gates" AND, OR and NOT. These three gates are thus said to form a "universal set" of gates.

A quantum circuit is similarly built out of logical quantum wires carrying qubits, and quantum gates acting on these qubits. Each wire corresponds to one of the *n* qubits. We assume each gate acts on either one or two wires. The possible physical transformations of quantum systems are unitary transformations[39], so such quantum gates can be described by a unitary matrix.

Any quantum computation can be expressed as a sequence of one-qubit and two-qubit unitary quantum gates. The most general one-qubit gate is described by a 2x2 unitary matrix :

(177) $\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}$

which maps |0> to $\alpha|0>+\beta|1>$ and |1> to $\gamma|0>+\delta|1>$. The condition of unitarity may be expressed in the form : the matrix above must be such that the transpose of its complex conjugate is equal to the inverse matrix.

[38] P. W. Shor, *Introduction to quantum algorithms*, quant-ph/0005003 (2000). Available online at http://xxx.adelaide.edu.au.

[39] See e.g. A. Messiah, *Quantum mechanics (volume 1)*, North-Holland, Amsterdam (1961) p.312.

§12. On the relative efficiencies of classical and quantum computing machines[40]

Apart from our observation that the state space in which a quantum computer lives is exponentially larger than its classical counterpart, our discussions have made no mention of the of the relative efficiencies of classical and quantum computing. Are quantum computers more efficient than classical computers? Even if quantum computers were only as powerful as classical computers, their study would nevertheless still be important. This is due to the fact, mentioned earlier, that at the current exponential rate of microelectronics miniaturization, it will not be long before integrated circuits will need to explicitly take quantum effects into consideration in order to function.

So, just how efficient are quantum computers relative to their classical counterparts? The answer depends on the type of computation one is trying to carry out.

(a) For factorising large numbers, quantum computers are *exponentially faster* than their classical counterparts. This may be of importance in encryption and code-cracking, as many contemporary encryption algorithms are based on the fact that factorisation is notoriously difficult for existing computers[41].

(b) For searching through a database to retrieve a certain piece of information at an unknown location, quantum computers are *quadratically*

[40] P. W. Shor, *Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer*, quant-ph/9508027 version 2 (1995, revised 1997); P. W. Shor, *Introduction to quantum algorithms*, quant-ph/0005003 (2000). Both papers available online at http://xxx.adelaide.edu.au. See also C. H. Bennett and D. P. Vincenzo (2000).

[41] Nowadays, with the best hardware and algorithm, it is barely possible to find the 65-digit prime factors of a 130-digit number in a few months. The difficulty escalates sharply as the number of digits increases - with the same hardware and algorithm, we would need about the age of the universe (10 billion years) to factor a 400-digit number. But, using Shor's algorithm (see previous footnote) and a quantum computer that could factor a 130-digit number in a month, we would be able to factor a 400-digit number in a few years. So, at least for this class of "hard" problem, the time needed to find a solution scales much more favorably with the size of the problem if we use a quantum computer rather than a classical computer.

faster than classical computers. This may be of importance in searching through vast databanks, such as future incarnations of the Internet.

(c) For the simulation of quantum-mechanical systems, quantum computers can give an exponential speed-up relative to classical computers. Interestingly, Feynman appears to be the first to have pointed this out[42].

(d) There are tasks doable on a quantum computer which have no counterpart classically. Thus, in some cases, a quantum computer can be infinitely more efficient than a classical computer. Quantum Cryptography provides an absolute secrecy of communication between parties that is impossible classically[43]. Also, for some games, winning strategies become possible with the use of quantum resources that are not available otherwise[44].

Unfortunately, we do not have time in this course to consider any quantum algorithms in detail. If you are interested, here are some references :

- Factorisation: P.W.Shor, *Polynomial-time algorithms for prime factorisation and discrete logarithms on a quantum computer* quant-ph/9508027 version 2 (1995, revised 1997), available online at http://xxx.adelaide.edu.au.

- Search : L.K.Grover, *Quantum mechanics helps in searching for a needle in a haystack*, Physical Review Letters **78** 325-328 (1997). Note that this technique is readily generalized to a number of other tasks.

- Fast Fourier Transform : P.W.Shor, *Introduction to quantum algorithms*, quant-ph/0005003 (2000), available online at http://xxx.adelaide.edu.au.

§13. Environmental interactions (decoherence) and the development of quantum fault-tolerant computation[45]

At the end of §10.2, we mentioned that the vast majority of quantum states are entangled to some degree. This observation is closely related to the vast dimensionality of the state space in which the quantum computer lives, and is crucially related to the massive parallelism of which quantum computers are capable.

However, entanglement is also responsible for a negative and potentially fatal effect on quantum computers. Entangling interactions with the environment are thought to be the main reason why the macroscopic world seems to behave classically rather than quantum-mechanically. Macroscopically different states (such as the charge states representing 0 and 1 in the memory cell of a classical computer) interact so strongly with the environment that information rapidly leaks out as to which state the memory cell is in. Therefore, even if it were possible to prepare the cell in a quantum-mechanical superposition of 0 and 1, it would rapidly evolve into a complex entangled state involving the environment, which from the viewpoint of the memory cell would appear as a statistical mixture (i.e. a mixed state) rather than a superposition of the classical values. The spontaneous decay of superpositions into mixtures is known as decoherence.

Entanglement with the environment is thus a major obstacle for quantum computation. To avoid having a quantum computation decohere into a probabilistic classical computation (which could just as well be done on a classical computer) it is necessary, while creating and maintaining entanglement among the computational degrees of freedom, to avoid entanglement between them and the environment.

[42] R. Feynman, *Simulating physics with computers*, International Journal of Theoretical Physics **21** 467-488 (1982) ; R. Feynman, *Quantum mechanical computers*, Foundations of Physics **16** 507-531 (1986).

[43] C. H. Bennett and G. Brassard, "Quantum Cryptography: Public Key Distribution and Coin Tossing", in *Proceedings of the IEEE International Conference on Computers, Systems and Signal Processing, Bangalore, India* (IEEE, New York, 1984), p. 175.

[44] D. A. Meyer, *Quantum strategies*, Physical Review Letters **82** 1052 (1999) ; J. Eisert, M. Wilkens and M. Lewenstein, *Quantum games and quantum strategies*, Physical Review Letters **83** 3077 (1999) ; L. Goldenberg, L. Vaidman and S. Wiesner, *Quantum gambling*, Physical Review Letters **82** 3356 (1999) ; A. M. Steane and W. van Dam, *Physicists triumph at "Guess my number"*, Physics Today **53** (2) 35-39 (2000).

[45] C. H. Bennett and D. P. Vincenzo (2000).

Until recently, it appeared that the feasible number of steps in a coherent quantum computation would necessarily be less than the ratio τ_d/τ_s of the decoherence time τ_d to the switching time τ_s characteristic of the elementary quantum systems used in the hardware of the quantum computer. Even if all other problems in the design of a practical quantum computer could be overcome, currently attainable values of τ_d/τ_s are not high enough to make quantum computers competitive with classical ones; also, the search for systems with ever-higher τ_d/τ_s might ultimately be blocked by fundamental properties of available atoms and nuclei. Apart from decoherence, it also appeared that individual gate operations would have to be made more and more precise the longer the computation.

This pessimism has largely been dispelled by the discovery of quantum fault-tolerant computation, the quantum analogue of existing methods on classical computers which allow arbitrarily-long classical computations to be reliably undertaken provided that the error probability per gate is less than some constant threshold. Because of quantum fault-tolerant computation, it appears that experimentalists need "only" build quantum hardware with a per-gate decoherence that is below some finite threshold in order for quantum computers to do arbitrarily complex calculations.

§14. Quantum software[46]

As we have already explained, a quantum computer is in essence a device which uses unitary operations to process information which is encoded in quantum states, such as the internal electronic states of individual atoms or the spin states of atomic nuclei. In addition to the formidable technological challenge in constructing large-scale quantum hardware, there is the challenge of creating what might be termed quantum software.

A quantum software program is a particular quantum state that enables a particular quantum computer to perform a specific task. If that quantum state is difficult or inconvenient to prepare, the user of a quantum computer might prefer to acquire the said quantum state from a vendor, rather than prepare it her/himself. The user's quantum computer hardware then acts on the quantum state according to the standard protocol.

Unfortunately for the user quantum software would be a consumable product, unavoidably damaged after a single use. Thus we might foresee the flourishing of a quantum software industry. A manufacturer can design a valuable quantum state, and use a special-purpose device to churn out multiple copies of the said state; these can then be tested to assure quality and then stored until needed. Consumers can then download then state for a fee and plug it into their own quantum computers to achieve improved performance. (In case you are wondering how quantum states might be distributed over the "Quantum Internet", see the discussion of "flying qubits" in the next section.)

The notion of quantum software also has some bearing on the previous section's discussion of quantum fault-tolerant computation. For each of the known quantum-error correction schemes, some of the gates in the fault-tolerant universal set are easy to perform, while others are hard. These "hard" gates would be most conveniently executed with quantum software that can be prepared ahead of time and then consumed during the operation of the gate. Thus we have the possibility that more affordable quantum computing machines (which lack the "hard" gates) might be marketed with only rudimentary hardware tools. Unable to execute a universal set of fault-tolerant gates with its hardware alone, it would be able to achieve that capability through quantum software suited to its coding scheme.

§15. Seven requirements for the experimental implementation of quantum computation[47]

We have reached a point in our discussions where we know enough about the fundamentals of Quantum Computing to be able to outline a list of criteria (due to D.P.Vincenzo (2000)) which it is believed must be met in any future experimental incarnation of a Quantum Computer.

[46] J. Preskill, *Plug-in quantum software*, Nature **402** 357-358 (1999).

[47] D. P. DiVincenzo, *The physical implementation of quantum computation,* quant-ph/0002077 version 2 (2000). Available online at http://xxx.adelaide.edu.au.

- *A scalable physical system with well-characterized qubits.* The qubits' physical parameters should be accurately known, including the internal Hamiltonian of the qubit (this determines the energy eigenstates of the qubit, which are often but not always taken as the |0> and |1> states), the presence of and couplings to other qubits, and the couplings to external fields that might be used to manipulate the state of the qubit.

- *The ability to initialise the state of the qubits to a simple fiducial state, such as |000...>.* This arises from the straightforward computing requirement that registers should be initialised to a known value before the start of a computation.

- *Long relevant decoherence times, much longer than the gate operation time.* We have already mentioned that decoherence is potentially very dangerous for quantum computing, because if it acts for very long, the capability of the quantum computer will reduce to that of a classical computer. We pointed out that quantum computing can be made fully fault tolerant if the decoherence time is sufficiently longer than the "clock" time of the quantum computer. Provided that quantum fault-tolerant computation is employed, contemporary estimates state that the decoherence time should be about 10,000 times the length of a single "tick" in order for error correction to be successful. This is a rather stringent condition. However, once the desired threshold has been reached, decoherence will no longer be an obstacle to quantum computations of arbitrary length (for reasons outlined in §29).

- *A universal set of quantum gates.* This needs little explanation. We have already pointed out that a quantum algorithm is typically a set of unitary transformations, each of which act on a small number of qubits (typically no more than three). A universal set of quantum gates is by definition sufficient to realise all possible unitary transformations, and is therefore a requirement for the construction of a "complete" quantum computer. This requirement can be relaxed, however, for reasons outlined in the last paragraph of §30.

- *A qubit-specific measurement capability.* The results of a quantum computation need to be read out, and this requires the ability to measure specific qubits!

- *The ability to interconvert stationary and flying qubits.* Stationary qubits are exactly that - stationary. An example might include various solid-state proposals for quantum computers. In contrast, "flying qubits" (can you picture their wings?) are qubits that are readily transferred from place to place. Almost all proposals assume that photon states, with the qubit either encoded in the polarization or the spatial wavefunction of the photon, will be the flying qubit of choice, and indeed the well-developed technology of light transmission through optical fibres provides a very promising system for the transmission of qubits. There exists the alternative possibility that electrons travelling through solids could provide a practical realization of flying qubits.

- *The ability faithfully to transmit flying qubits between specified locations.* In order for quantum computers to communicate with one another over the Quantum Internet, quantum information will need to be transferred from the stationary qubits of a given quantum computer source, into flying qubits which can faithfully travel over an appropriate network (e.g. optical fibres), and then be read into a given quantum computer target.

§16. Two examples of the state-of-the-art in experimental studies of quantum computation

§16.1 A scalable method for creating entangled states[48]

As we have seen, the creation of entangled states is of fundamental importance to the realization of quantum computers. Very recently (March 2000), the experimental quantum entanglement of four particles was achieved by Sackett and co-workers.

[48] Rainer Blatt, *Push-button entanglement*, Nature **404** 231-232 (2000) ; C. A. Sackett, D. Kielpinski, B. E. King, C. Langer, V. Meyer, C. J. Myatt, M. Rowe, Q. A. Turchette, W. M. Itano, D. J. Wineland and C. Monroe, *Experimental entanglement of four particles*, Nature **404** 256-259 (2000).

Before these experiments, only two or three particles had been entangled. However, this latest work on four-particle entanglement is not just an incremental improvement. Unlike the previous work, this latest work is applicable to creating much larger numbers of entangled particles, and therefore constitutes a *scalable method for creating entangled states*.

The reason that previous experimental methods for creating entanglement were not scalable is because they relied on selection of data from certain random processes. All such methods suffer from the problem that the probability of randomly generating the appropriate conditions decreases exponentially with the number of particles one is attempting to entangle. For example, in a recent experiment which demonstrated three-photon entanglement, two-photon entangled states could be generated and detected at the rate of about 1000 per second ; three-photon entangled states at a rate of about 30 per hour ; and four-photon states at an extrapolated rate of several per year.

The scalable procedure used to achieve four-particle entanglement is roughly as follows. Four spin-half charged particles are confined in a harmonic trap. Using laser cooling[49] and other techniques, the particles are initially prepared in their spin-down state and in the ground state of their collective motion : $|\psi\rangle=|\downarrow\downarrow\downarrow\downarrow 0\rangle$. A suitable set of laser pulses then results in the deterministic (as opposed to probabilistic) creation of a maximally-entangled state of four particles.

§16.2 Interconversion of stationary and flying qubits[50]

Workers have proposed a marriage of techniques from photon-fibre systems and trapped-atom (or trapped-ion) systems. In these schemes, a "standing" qubit encoded in a state of an atom is mapped by an appropriate laser pulse into the same qubit of the photon state of the surrounding electromagnetic cavity. The quantum information can from there become a flying qubit by leaking out into the propagating mode in free space or in an optical fibre.

[49] This refers to the remarkable technique whereby the shining of appropriately-configured laser beams onto a cloud of atoms results in the cooling of those atoms to mere nano-Kelvins above absolute zero.

[50] C. H. Bennett and D. P. DiVincenzo, *Quantum computing - towards an engineering era?*, Nature **377** 389 (1995) ; S. J. van Enk, J. I. Cirac and P. Zoller, *Ideal quantum communication over noisy channels: a quantum optical implementation*, Physical Review Letters **78** 4293-4296 (1997) ; S. J. van Enk, H. J. Kimble, J. I. Cirac and P. Zoller, *Quantum communication with dark photons*, Physical Review A **59** 2659-2664 (1999).

CHAPTER 3
QUANTUM THEORY OF RADIATION

§17. Introduction and overview

In this last section of the course, we show how to *quantise* the electromagnetic field by treating the electric and magnetic fields as appropriately-defined operators. We give some application of these ideas to the problem of so-called "phase operators" in photon optics.

§18. The electromagnetic field in the absence of charges

§18.1 The classical electromagnetic field[51]

Classical electromagnetic theory is summed up in Maxwell's equations:

(178a) $\nabla \bullet \vec{E} = \rho$ ... Maxwell I

(178b) $\nabla \times \vec{B} = c^{-1}\vec{j} + c^{-1}\partial_t \vec{E}$ ... Maxwell II

(178c) $\nabla \bullet \vec{B} = 0$ ... Maxwell III

(178d) $\nabla \times \vec{E} = -c^{-1}\partial_t \vec{B}$ ... Maxwell IV

where we are working in rationalized Lorentz-Heaviside units, $\vec{E}$ and $\vec{B}$ are the electric and magnetic fields, ρ and $\vec{j}$ are the charge and current densities, c is the speed of light and t is time.

Claim: From the second pair of Maxwell equations follows the existence of scalar and vector potentials $\varphi(\vec{x},t)$ and $\vec{A}(\vec{x},t)$, defined by:

[51] F. Mandl and G. Shaw, *Quantum Field Theory*, Revised edition, John Wiley & Sons, Chichester (1993), section 1.2.1.

(179a) $\vec{B} = \nabla \times \vec{A}$

(179b) $\vec{E} = -\nabla\varphi - c^{-1}\partial_t \vec{A}$.

►*Exercise 25: Prove equations (179). Hence, or otherwise, show that the second pair of Maxwell equations is automatically satisfied when one works in terms of the potentials* φ *and* $\vec{A}$*, rather than in terms of the fields* $\vec{E}$ *and* $\vec{B}$.

Claim: For an arbitrary function $f(\vec{x}, t)$, the transformation:

(180) $$\begin{cases} \varphi \to \varphi' = \varphi + c^{-1}\partial_t f \\ \vec{A} \to \vec{A}' = \vec{A} - \nabla f \end{cases}$$

leaves the electric and magnetic fields unaltered. This transformation is known as a "gauge transformation of the second kind".

►*Exercise 26: Prove that the gauge transformations of the second kind leave the electric and magnetic fields unaltered. Comment briefly on the consequent non-uniqueness of the scalar and vector potentials for the electromagnetic field. Does this non-uniqueness, known as "gauge freedom", have any physical significance in the context of classical electrodynamics?*

We have seen, in exercise (25), that the second pair of Maxwell equations is automatically satisfied when one works in terms of the potentials φ and $\vec{A}$, rather than in terms of the fields $\vec{E}$ and $\vec{B}$. What about the first pair of Maxwell equations? Well, substitution of (179b) into Maxwell I gives:

(181) $$\begin{aligned} &\nabla \bullet \left(-\nabla\varphi - c^{-1}\partial_t \vec{A}\right) = \rho \\ &-\nabla^2\varphi - c^{-1}\partial_t \nabla \bullet \vec{A} = \rho \end{aligned}$$

while substitution of both (179a) and (179b) into Maxwell II gives:

(182) $\nabla \times \nabla \times \vec{A} = c^{-1}\vec{j} + c^{-1}\partial_t\left(-\nabla\varphi - c^{-1}\partial_t \vec{A}\right)$.

Next, recall the following identity from vector analysis:

(183) $\nabla \times \nabla \times \vec{A} = \nabla\nabla \bullet \vec{A} - \nabla^2 \vec{A}$,

allowing us to transform (182) into:

(184) $$\begin{aligned} &\nabla\nabla \bullet \vec{A} - \nabla^2 \vec{A} = c^{-1}\vec{j} - c^{-1}\nabla\partial_t\varphi - c^{-2}\partial_t^2 \vec{A} \\ &\left(c^{-2}\partial_t^2 - \nabla^2\right)\vec{A} + \nabla\left(\nabla \bullet \vec{A} + c^{-1}\partial_t\varphi\right) = c^{-1}\vec{j} \end{aligned}$$.

Equations (181) and (184) are complicated. To make them easier to work with, we can make use of the gauge freedom discovered in the last exercise. This freedom allows us to demand that the vector potential be divergence free:

(185) $\nabla \bullet \vec{A} = 0$.

Such a "choice of gauge" is known as the "Coulomb gauge" or the "radiation gauge".

►*Exercise 27: Justify the claim that one can always choose a gauge for the potentials such that (185) holds true.*

We now show how the Coulomb gauge leads to a dramatic simplification of (181) and (184). In the absence of both charges and currents (where ρ and $\vec{j}$ are zero), equation (181) reduces to:

(186) $\nabla^2\varphi = 0$,

which has the solution

(187) $\varphi = 0$.

Thus the electric potential drops out of the picture altogether! If we substitute (187) into (184), adopt the Coulomb gauge (185), and set the current $\vec{j}$ to zero, we arrive at the following *wave equation*:

(188) $\left(c^{-2}\partial_t^2 - \nabla^2\right)\vec{A} = 0$.

In the Coulomb gauge, equations (179) become:

(189a) $\vec{B} = \nabla \times \vec{A}$ (189b) $\vec{E} = -c^{-1}\partial_t \vec{A}$.

Therefore, in the Coulomb gauge and in the absence of both charges and currents, the electric and magnetic fields may both be determined from the vector potential $\vec{A}$.

With reference to the Coulomb condition (185), we note that a vector field with vanishing divergence is often called a "transverse vector field", since for a polarized plane wave:

(190) $\vec{A}(\vec{x}, t) = \vec{A}_0 \exp\left(i\left(\vec{k} \bullet \vec{x} - \omega t\right)\right)$,

the Coulomb condition implies:

(191) $\vec{k} \bullet \vec{A} = 0$.

We shall have need for this result a little later.

►*Exercise 28: Prove (191), and then briefly interpret your result.*

As a prelude to the quantisation of the classical electromagnetic field, we consider the radiation to exist inside a large cubic enclosure[52], of side-length L and volume $V = L^3$. We impose periodic boundary conditions on the vector potential $\vec{A}$ at the surfaces of the cube, i.e. we demand that:

(192) $$\vec{A}(x,y,z) = \vec{A}(x + n_1 L, y + n_2 L, z + n_3 L), \quad n_1, n_2, n_3 = 0, \pm 1, \pm 2 \text{ etc.}$$

With such periodic boundary conditions, the functions:

(193) $$\frac{1}{\sqrt{V}} \vec{\varepsilon}_r(\vec{k}) \exp(i\vec{k} \bullet \vec{x}), \quad r = 1,2$$

form a complete set of transverse ortho-normal vector fields. Here, the periodic boundary conditions imply that the wave-vectors $\vec{k}$ must be of the form:

(194) $$\vec{k} = \frac{2\pi}{L}(n_1, n_2, n_3), \; n_1, n_2, n_3 = 0, \pm 1, \pm 2 \text{ etc.}$$

The quantities $\vec{\varepsilon}_1(\vec{k})$ and $\vec{\varepsilon}_2(\vec{k})$ are two mutually perpendicular real unit vectors, both of which are orthogonal to $\vec{k}$:

(195) $\vec{\varepsilon}_r(\vec{k}) \bullet \vec{\varepsilon}_s(\vec{k}) = \delta_{rs}, \; \vec{\varepsilon}_r(\vec{k}) \bullet \vec{k} = 0, \; r,s = 1,2$.

The second of these two conditions follows from (191).

►*Exercise 29: Prove the orthonormality of the vector fields in (193). Show that the wave-vectors in (194) ensure that the vector fields in (193) satisfy the required periodic boundary conditions. Give an interpretation of the vectors* $\vec{\varepsilon}_1(\vec{k})$ *and* $\vec{\varepsilon}_2(\vec{k})$.

We can now expand our vector potential as a Fourier series, i.e. *at each instant of time we can consider the vector potential to be a weighted sum of plane waves* with the various wavenumbers allowed by (194):

(196)
$$\vec{A}(\vec{x},t) = \sum_{\vec{k}} \sum_r \sqrt{\frac{\hbar c^2}{2V\omega_k}} \vec{\varepsilon}_r(\vec{k}) \left(a_r(\vec{k},t) e^{i\vec{k} \bullet \vec{x}} + a_r^*(\vec{k},t) e^{-i\vec{k} \bullet \vec{x}} \right),$$

where $\omega_k = c|\vec{k}|$. The summations with respect to r and $\vec{k}$ are over both polarization states $r = 1,2$ (for each $\vec{k}$) and over all allowed wave-numbers $\vec{k}$ (cf. (194)). The factor under the square root sign has been introduced for later convenience. The form of the Fourier series in (196) ensures that the vector potential is real.

Since we are working in the Coulomb gauge, the vector potential given by (196) must satisfy the wave equation (188). If you substitute (196) into the wave equation (188), you will find that:

(197) $\partial_t^2 a_r(\vec{k},t) = -\omega_k^2 a_r(\vec{k},t)$.

►*Exercise 30: Derive equation (197).*

Equations (197) are a set of *uncoupled harmonic oscillator equations* for the Fourier coefficients $a_r(\vec{k},t)$ of the vector potential, whose solution can be written down immediately:

(198) $a_r(\vec{k},t) = a_r(\vec{k}) \exp(-i\omega_k t)$,

allowing us to re-write (196) as:

(199)
$$\vec{A}(\vec{x},t) = \sum_{\vec{k}} \sum_r \sqrt{\frac{\hbar c^2}{2V\omega_k}} \vec{\varepsilon}_r(\vec{k}) \left(a_r(\vec{k}) e^{i\vec{k} \bullet \vec{x}} e^{-i\omega_k t} + a_r^*(\vec{k}) e^{-i\vec{k} \bullet \vec{x}} e^{i\omega_k t} \right)$$

Thus we have an expansion of the vector potential $\vec{A}(\vec{x},t)$ as a linear combination of the monochromatic plane-wave "modes" $\vec{\varepsilon}_r(\vec{k}) e^{i\vec{k} \bullet \vec{x}} e^{-i\omega_k t}$ and $\vec{\varepsilon}_r(\vec{k}) e^{-i\vec{k} \bullet \vec{x}} e^{i\omega_k t}$. The mode $\vec{\varepsilon}_r(\vec{k}) e^{i\vec{k} \bullet \vec{x}} e^{-i\omega_k t}$ may

[52] Those who dislike the introduction of a fictitious box when quantising the electromagnetic field might like to look up a treatment which avoids this artifice. See e.g. A. Messiah, *Quantum Mechanics (volume 2)*, North-Holland, Amsterdam (1961), p. 974.

be thought of as a *polarized monochromatic plane wave*, which is linearly polarized along the direction $\vec{\varepsilon}_r$; the polarization vector $\vec{\varepsilon}_r$ is perpendicular to the vector $\vec{k}$ which points in the direction of propagation of the plane wave; $k = 2\pi/\lambda = |\vec{k}|$ and so the wavelength λ of the plane wave is equal to $2\pi/|\vec{k}|$; the momentum $\vec{p}$ of the plane wave mode is $\vec{p} = \hbar\vec{k}$. A similar interpretation exists for the mode $\vec{\varepsilon}_r(\vec{k})e^{-i\vec{k}\bullet\vec{x}}e^{i\omega_k t}$.

§18.2 Two interludes

In this interlude we develop two pre-requisite pieces of knowledge which, once assimilated, will allow us to continue with our development of the quantum theory of the photon field.

(a) The first interlude reviews the theory of the harmonic oscillator from the point of view of the so-called *creation and destruction operators*. Such a discussion is relevant in light of our discovery, in equation (197), that the Fourier coefficients of the vector potential satisfy a set of uncoupled harmonic oscillator equations.

(b) The second interlude concerns the so-called Schrödinger and Heisenberg "pictures". In the former picture, the wavefunctions are considered to vary with time while the operators are time independent. In the latter picture, the operators are considered to vary with time while the wavefunction is time independent.

§18.2.1 Interlude #1: The harmonic oscillator, revisited[53]

We review the theory of the harmonic oscillator from the point of view of the so-called *creation and destruction operators*. This method is due to Dirac.

The Hamiltonian $\hat{H}$ for the harmonic oscillator is given by a sum of kinetic and potential energies:

(200) $\hat{H} = \frac{\hat{p}^2}{2m} + \frac{1}{2}m\omega^2\hat{q}^2$,

where $\hat{p}$ and $\hat{q}$ are momentum and position operators[54] satisfying the following commutation relation:

(201) $[\hat{q},\hat{p}] = i\hbar$,

m is the mass of the particle, and ω is the angular frequency (of the oscillator, when considered classically). We now introduce the operators:

(202) $\hat{a} \equiv \frac{m\omega\hat{q} + i\hat{p}}{\sqrt{2\hbar m\omega}}$, $\hat{a}^\dagger = \frac{m\omega\hat{q} - i\hat{p}}{\sqrt{2\hbar m\omega}}$.

Equations (201) and (202) together imply that:

(203) $[\hat{a},\hat{a}^\dagger] = 1$.

►*Exercise 31: Derive equation (203).*

The Hamiltonian, expressed in terms of $\hat{a}$ and $\hat{a}^\dagger$, becomes:

(204) $\hat{H} = \hbar\omega\left(\hat{a}^\dagger\hat{a} + \frac{1}{2}\right)$

►*Exercise 32: Derive equation (204).*

With a view to interpreting this result, let us consider the operator:

(205) $\hat{N} \equiv \hat{a}^\dagger\hat{a}$.

The expectation value of $\hat{N}$ is non-negative, as is readily shown:

(206) $\langle\Psi|\hat{N}|\Psi\rangle = \langle\Psi|\hat{a}^\dagger\hat{a}|\Psi\rangle = \langle\hat{a}\Psi\,|\,\hat{a}\Psi\rangle \geq 0$.

Let $|\beta\rangle$ be an eigenvector of $\hat{N}$ with eigenvalue β:

(207) $\hat{N}|\beta\rangle = \beta|\beta\rangle$.

Hence:

(208) $\langle\beta|\hat{N}|\beta\rangle = \langle\beta|\beta|\beta\rangle = \beta\langle\beta\,|\,\beta\rangle \geq 0$,

where the last step follows from (206). Since $\langle\beta\,|\,\beta\rangle \geq 0$, the inequality at the end of (208) implies

[53] F. Mandl and G. Shaw, *Quantum Field Theory*, Revised edition, John Wiley & Sons, Chichester (1993), section 1.2.2. See also A. Messiah, *Quantum Mechanics (volume 1)*, North-Holland, Amsterdam (1961), chapter XII.

[54] For the remainder of the course, a "hat" (caret) over a quantity indicates it to be an operator.

that $\beta \geq 0$. Hence $\hat{N}$ possesses a lowest non-negative eigenvalue, which we denote by β_0:

(209) $\beta_0 \geq 0$.

Put this result to one side for the moment. Now, using equations (203) and (207), it can be shown that:

(210a) $\hat{N}\hat{a}|\beta\rangle = (\beta - 1)\hat{a}|\beta\rangle$

(210b) $\hat{N}\hat{a}^\dagger|\beta\rangle = (\beta + 1)\hat{a}^\dagger|\beta\rangle$

►*Exercise 33: Derive equations (210).*

We see from equations (210) that both $\hat{a}|\beta\rangle$ and $\hat{a}^\dagger|\beta\rangle$ are eigenfunctions of $\hat{N}$, with eigenvalues of $\beta - 1$ and $\beta + 1$ respectively.

Next, let $\beta = \beta_0$ in (210a), which gives:

(211) $\hat{N}\hat{a}|\beta_0\rangle = (\beta_0 - 1)\hat{a}|\beta_0\rangle$.

Suppose that $\hat{a}|\beta_0\rangle \neq 0$; as a consequence of this assumption, (211) implies that $\hat{a}|\beta_0\rangle$ is an eigenfunction of $\hat{N}$ with eigenvalue $\beta_0 - 1$. This eigenvalue is smaller than the smallest eigenvalue β_0. Our assumption has led to a logical contradiction, therefore the initial assumption is incorrect, therefore:

(212) $\hat{a}|\beta_0\rangle = 0$.

Furthermore, from (205) and (207) we have:

(213) $\hat{a}^\dagger\{\hat{a}|\beta_0\rangle\} = \beta_0|\beta_0\rangle$.

The quantity in braces is zero, on account of (212), and so the left side of (213) is zero. The ket $|\beta_0\rangle$ on the right side of (213) is non-zero, therefore:

(214) $\beta_0 = 0$.

It follows from (210b) that the eigenvalues n of $\hat{N}$ are the integers:

(215) $n = 0,1,2,3...$

We denote the eigenvectors of eigenvalue n by $|n\rangle$, and so (207) becomes:

(216) $\hat{N}|n\rangle = n|n\rangle$, $n = 0,1,2,3...$

We demand that the states $|n\rangle$ be normalised:

(217) $\langle n \mid n\rangle = 1$.

It follows that the states $|n \pm 1\rangle$ defined by:

(218a) $\hat{a}|n\rangle = \sqrt{n}|n-1\rangle$

(218b) $\hat{a}^\dagger|n\rangle = \sqrt{n+1}|n+1\rangle$

are also normalised.

►*Exercise 34: Derive equations (218). Show that the states, constructed using these equations, are normalised.*

The normalised eigenfunctions $|n\rangle$ of $\hat{N}$ can be constructed from the so-called "vacuum state" $|0\rangle$ in the following manner:

(219) $|n\rangle = \frac{(\hat{a}^\dagger)^n}{\sqrt{n!}}|0\rangle$, $n = 0,1,2,3....$

►*Exercise 35: Derive equation (219).*

From (204) and (216), we see that the states $|n\rangle$ are eigenfunctions of the harmonic oscillator Hamiltonian, with energy eigenvalues:

(220) $E_n = \hbar\omega\left(n + \frac{1}{2}\right)$, $n = 0,1,2,3...$

The operators $\hat{a}$ and $\hat{a}^\dagger$ are called *destruction* and *creation* operators respectively, on account of equations (218)[55]. Bearing (220) in mind, we see that the ket $|n\rangle$ *represents a state with n quanta of energy;* this state is known as the "number state" or "Fock state". We can therefore meaningfully call $\hat{N} \equiv \hat{a}^\dagger\hat{a}$ the "number operator", since its eigenvalue tells us the number n of quanta in a given "number state" $|n\rangle$ (cf. (205) and (216)).

The operator $\hat{a}$ acts on $|n\rangle$ to give $\sqrt{n}|n-1\rangle$, thus

[55] In the literature, the following terms are used interchangeably: "destruction operator", "absorption operator", "annihilation operator".

destroying a quantum of energy from the harmonic oscillator (cf. (218a)); similarly, $\hat{a}^\dagger$ acts on $|n\rangle$ to give $\sqrt{n+1}|n+1\rangle$, thus creating a quantum of energy (cf. (218b)). Since the operators $\hat{a}$ and $\hat{a}^\dagger$ take us up and down the scale of energy, they are sometimes collectively referred to as "ladder operators".

§18.2.2 Interlude #2: The Schrödinger and Heisenberg pictures[56]

This second interlude concerns the so-called Schrödinger and Heisenberg "pictures". In the former picture, the wavefunctions are considered to vary with time while the operators are time independent. In the latter picture, the operators are considered to vary with time while the wavefunction is time independent. The two pictures are thereby two different ways of describing the time development of a given quantum system.

In the Schrödinger picture, which is the one that you are familiar with, the time dependence is carried by the wavefunction (i.e. the wavefunction varies with time). We denote this time-dependent wavefunction by $|\Psi(t)\rangle_S$, where the "S" subscript indicates that we are working in the Schrödinger picture. The Schrödinger-picture wavefunction satisfies the Schrödinger equation:

$$(222)\quad i\hbar\frac{d}{dt}|\Psi(t)\rangle_S = \hat{H}|\Psi(t)\rangle_S .$$

The Schrödinger equation can be formally solved by writing down:

$$(223)\quad |\Psi(t)\rangle_S = \hat{U}(t,t_0)|\Psi(t=t_0)\rangle_S ,$$

where the "time-evolution operator" is:

$$(224)\quad \hat{U}(t,t_0) \equiv \exp\left(-i(t-t_0)\hat{H}/\hbar\right) \equiv \hat{U} .$$

►*Exercise 36: Show that (223) is indeed a solution of (222). Interpret the time-evolution operator* $\hat{U}$*, and show that it is unitary, i.e.* $\hat{U}^\dagger\hat{U} = \hat{U}\hat{U}^\dagger = \hat{I}$*, where* $\hat{I}$ *is the unit operator.*

Evidently, the inverse of the time-evolution operator can be written down by interchanging t and t_0:

$$(225)\quad \hat{U}^{-1} = \hat{U}(t_0,t) \equiv \exp\left(+i(t-t_0)\hat{H}/\hbar\right) \equiv \hat{U}^\dagger .$$

By means of $\hat{U}$ we can carry out a unitary transformation of state vectors and operators from the Schrödinger picture to the Heisenberg picture, according to:

$$(226a)\quad |\Psi\rangle_H \equiv \hat{U}^\dagger|\Psi(t)\rangle_S = |\Psi(t=t_0)\rangle_S .$$

$$(226b)\quad \hat{O}^H(t) = \hat{U}^\dagger\hat{O}^S\hat{U} .$$

Definition (226a) is straightforward, for it says that the Heisenberg representation $|\Psi\rangle_H$ of the state vector is, by definition, obtained by applying the inverse-time-evolution operator $\hat{U}^\dagger$ to $|\Psi(t)\rangle_S$ in order to "wind the clock back" to yield $|\Psi(t=t_0)\rangle_S$. Stated differently, the Heisenberg state vector is equal to the Schrödinger state vector at time $t=t_0$; *the state vector therefore has no time dependence in the Heisenberg picture*. Rather, it is the *operators which carry the time dependence in the Heisenberg picture*, as given by (226b).

This argument motivates why the operators carry time dependence in the Heisenberg picture, but it does not serve to *derive* (226b). We now show that (226b) is a consequence of the definition (226a). Suppose that one has some operator $\hat{O}^S$ and two wavefunctions $|\Psi\rangle_S$ and $|\Phi\rangle_S$, where the operator O^S is such that:

$$(227)\quad \hat{O}^S|\Psi\rangle_S = |\Phi\rangle_S$$

Apply $\hat{U}^\dagger$ to both sides of this equation, to give:

$$(228)\quad \hat{U}^\dagger\hat{O}^S|\Psi\rangle_S = \hat{U}^\dagger|\Phi\rangle_S$$

Since $\hat{U}\hat{U}^\dagger = 1$ (see exercise 36), we can re-write (228) as:

$$(229)\quad \hat{U}^\dagger\hat{O}^S\hat{U}\left\{\hat{U}^\dagger|\Psi\rangle_S\right\} = \left\{\hat{U}^\dagger|\Phi\rangle_S\right\}.$$

[56] F. Mandl and G. Shaw, *Quantum Field Theory*, Revised edition, John Wiley & Sons, Chichester (1993), section 1.5. I also recommend the discussion given in section 5.8 of B. H. Bransden and C. J. Jochain, *Introduction to Quantum Mechanics*, Longman, Essex (1989).

The terms in braces are the Heisenberg representations of the state vectors Ψ and Φ (cf. (226a)), so:

$$\hat{U}^\dagger \hat{O}^S \hat{U} \left|\Psi\right\rangle_H = \left|\Phi\right\rangle_H . \quad (230)$$

The operator appearing on the left side of this expression is equal to $\hat{O}^H$, thus completing our proof.

The transformation between the Schrödinger and Heisenberg pictures leaves invariant a Hamiltonian which does not carry explicit time dependence:

$$\hat{H}^H = \hat{H}^S \equiv \hat{H} . \quad (231)$$

►*Exercise 37: Show that the transformation between Schrödinger and Heisenberg pictures leaves invariant both the commutation relation* $[\hat{A},\hat{B}] = \hat{C}$*, and the matrix element* $\left\langle \Psi_A \left| \hat{O} \right| \Psi_B \right\rangle$. *Here,* $\hat{A},\hat{B},\hat{C},\hat{O}$ *are operators and* $\left|\Psi_A\right\rangle, \left|\Psi_B\right\rangle$ *are state vectors.*

We close this section by deriving the Heisenberg equation of motion which governs the time evolution of operators in the Heisenberg picture. To start off the derivation, take the time derivative of (226b), making use of the product rule for the derivative of a product of three objects:

$$\begin{aligned} \frac{d}{dt}\hat{O}^H(t) &= \frac{d}{dt}\left(\hat{U}^\dagger \hat{O}^S \hat{U}\right) \\ &= \left(\frac{\partial \hat{U}^\dagger}{\partial t}\right)\hat{O}^S\hat{U} + \hat{U}^\dagger\left(\frac{\partial \hat{O}^S}{\partial t}\right)\hat{U} + \hat{U}^\dagger\hat{O}^S\left(\frac{\partial \hat{U}}{\partial t}\right). \end{aligned} \quad (232)$$

Assuming that the operator $\hat{O}^S$ has no explicit time dependence, the middle term on the right-hand side of (232) is zero. By taking the time derivative of (224), we find that $\partial\hat{U}/\partial t = \left(-i\hat{H}/\hbar\right)\hat{U}$. Similarly, the time derivative of the Hermitian conjugate of (224) shows that $\partial\hat{U}^\dagger/\partial t = \left(i\hat{H}/\hbar\right)\hat{U}^\dagger$. Thus (232) becomes:

$$\begin{aligned} \frac{d}{dt}\hat{O}^H(t) &= \left(\frac{i\hat{H}}{\hbar}\hat{U}^\dagger\right)\hat{O}^S\hat{U} - \hat{U}^\dagger\hat{O}^S\frac{i\hat{H}}{\hbar}\hat{U} \\ &= \frac{i}{\hbar}\left(\hat{H}\hat{U}^\dagger\hat{O}^S\hat{U} - \hat{U}^\dagger\hat{O}^S\hat{H}\hat{U}\right) \\ &= \frac{i}{\hbar}\left(\hat{H}\hat{U}^\dagger\hat{O}^S\hat{U} - \hat{U}^\dagger\hat{O}^S\left\{\hat{U}\hat{U}^\dagger\right\}\hat{H}\hat{U}\right) \\ &= \frac{i}{\hbar}\left(\hat{H}\hat{O}^H(t) - \hat{O}^H(t)\hat{U}^\dagger\hat{H}\hat{U}\right) \\ &= \frac{i}{\hbar}\left(\hat{H}\hat{O}^H(t) - \hat{O}^H(t)\hat{H}\right) \\ &= \frac{i}{\hbar}[\hat{H},\hat{O}^H(t)] \end{aligned} \quad (233)$$

Therefore we obtain the Heisenberg equation of motion:

$$i\hbar\frac{d}{dt}\hat{O}^H(t) = [\hat{O}^H(t),\hat{H}] . \quad (234)$$

► *Exercise 38: Show that, when we take* $\hat{O}^H$ *to be the creation operator associated with the harmonic oscillator, then the Heisenberg equation of motion implies that:*

$$\frac{d^2}{dt^2}\hat{a}^H(t) = -\omega^2\hat{a}^H(t), \quad (235)$$

which has the solution:

$$\hat{a}(t) = \exp(-i\omega t) . \quad (236)$$

§18.3 The quantised radiation field[57]

Now that we have assimilated the pre-requisites of §18.2, we can further the discussion which culminated in equation (199).

Equation (199) gave a Fourier-series expansion of the vector potential $\vec{A}$ associated with the classical electromagnetic field; under the Coulomb gauge (185), we saw that both the electric and magnetic fields may be derived from this vector potential according to equations (189). We also saw that each Fourier coefficient a_r, which appears in (199), independently satisfies the harmonic oscillator equation (197). But *we know how to quantise the harmonic oscillator*, using the ideas in §18.2.1.

[57] F. Mandl and G. Shaw, *Quantum Field Theory*, Revised edition, John Wiley & Sons, Chichester (1993), section 1.2.3.

Therefore we can quantise (199). The resulting *operator* for the vector potential of the *quantised electromagnetic field* will evidently be time-dependent, which is why we studied the Heisenberg picture in §18.2.2. This chain of logic, due to Dirac, will be followed through in the present section.

In view of the connections we have just made, the vector potential for the *quantised* electromagnetic field is:

(237)

$$\hat{\vec{A}}(\vec{x},t)=\sum_{\vec{k}}\sum_{r}\sqrt{\frac{\hbar c^2}{2V\omega_k}}\,\vec{\varepsilon}_r(\vec{k})\begin{pmatrix}\hat{a}_r(\vec{k})e^{i\vec{k}\bullet\vec{x}}e^{-i\omega_k t}\\ +\hat{a}_r^{\dagger}(\vec{k})e^{-i\vec{k}\bullet\vec{x}}e^{i\omega_k t}\end{pmatrix},$$

which is just (199) with the Fourier coefficients a and a^* replaced by the *operators* $\hat{a}$ and $\hat{a}^{\dagger}$. The operator $\hat{a}_r(\vec{k})$ may be thought of as the operator which destroys the photon mode $\vec{\varepsilon}_r(\vec{k})e^{i\vec{k}\bullet\vec{x}}e^{-i\omega_k t}$; similarly, $\hat{a}_r^{\dagger}(\vec{k})$ may be thought of as creating the photon mode $\vec{\varepsilon}_r(\vec{k})e^{-i\vec{k}\bullet\vec{x}}e^{i\omega_k t}$.

Despite the formal similarity of formulae (199) and (237), their meaning is very different. The vector potential of (199) is a classical vector field which has three components at each point in space-time. The vector potential of (237) is a Heisenberg operator that acts on state vectors.

We now have the vector potential for the quantised electromagnetic field. Since we are working in the Coulomb gauge, we substitute (237) into (189) to obtain the electric and magnetic field operators for the quantised electromagnetic field. The result is:

(238a)

$$\hat{\vec{E}}(\vec{x},t)=i\sum_{\vec{k}}\sum_{r}\sqrt{\frac{\hbar\omega_k}{2V}}\,\vec{\varepsilon}_r(\vec{k})\begin{pmatrix}\hat{a}_r(\vec{k})e^{i\vec{k}\bullet\vec{x}}e^{-i\omega_k t}\\ -\hat{a}_r^{\dagger}(\vec{k})e^{-i\vec{k}\bullet\vec{x}}e^{i\omega_k t}\end{pmatrix}$$

(238b)

$$\hat{\vec{B}}(\vec{x},t)=i\sum_{\vec{k}}\sum_{r}\sqrt{\frac{\hbar c^2}{2V\omega_k}}\,(\vec{k}\times\vec{\varepsilon}_r(\vec{k}))\begin{pmatrix}\hat{a}_r(\vec{k})e^{i\vec{k}\bullet\vec{x}}e^{-i\omega_k t}\\ -\hat{a}_r^{\dagger}(\vec{k})e^{-i\vec{k}\bullet\vec{x}}e^{i\omega_k t}\end{pmatrix}$$

►*Exercise 39: Prove equations (238a) and (238b). For the proof of the latter equation, you may need the identity:* $\nabla\times(A\vec{B})=\nabla A\times\vec{B}+A\nabla\times\vec{B}$. *Briefly interpret your results.*

Equations (237) and (238) give the operators for the vector potential, electric field, and magnetic field in the quantum theory of radiation. These Heisenberg operators act on state vectors of the radiation field. Since the quantised radiation field may be thought of as a collection of independent harmonic oscillators, one for each photon mode of allowed wavevector and polarization, we may appropriately generalize the "number states" of §18.2.1. This generalization will occupy us for the remainder of this section.

The creation and annihilation operators for *different* modes satisfy the so-called boson commutation relations:

(239a) $[\hat{a}_r(\vec{k}),\hat{a}_s^{\dagger}(\vec{k}')]=\delta_{rs}\delta_{\vec{k}\vec{k}'}$

(239b) $[\hat{a}_r(\vec{k}),\hat{a}_s(\vec{k}')]=[\hat{a}_r^{\dagger}(\vec{k}),\hat{a}_s^{\dagger}(\vec{k}')]=0$,

where the Kronecker delta δ_{rs} is equal to one if both subscripts are equal, and zero otherwise.

The total energy of the quantised radiation field is obtained by adding up the energies of each of the independent oscillators, and so equation (204) generalises to:

(240) $\hat{H}_{rad}=\sum_{\vec{k}}\sum_{r}\hbar\omega_k\left(\hat{a}_r^{\dagger}(\vec{k})\hat{a}_r(\vec{k})+\tfrac{1}{2}\right)$,

where $\omega_k=c|\vec{k}|$.

The number operator of (205) becomes:

(241) $\hat{N}_r(\vec{k})=\hat{a}_r^{\dagger}(\vec{k})\hat{a}_r(\vec{k})$,

which yields the number $n_r(\vec{k})=0,1,2,3,...$ of photons in a mode of given wavenumber $\vec{k}$ and polarization $\vec{\varepsilon}_r(\vec{k})$. The eigenfunctions of (241) are of the form:

(242) $$\left|n_r(\vec{k})\right\rangle=\frac{\left(\hat{a}_r^{\dagger}(\vec{k})\right)^{n_r(\vec{k})}}{\sqrt{n_r(\vec{k})!}}\left|0\right\rangle,$$

which is the generalization of (219). The eigenfunctions of the radiation Hamitonian (240) are the products of such states, i.e.:

(243) $$\left|\cdots,n_r(\vec{k}),\cdots\right\rangle=\prod_{\vec{k}_i}\prod_{r_i}\left|n_{r_i}(\vec{k}_i)\right\rangle.$$

The state $\left|\cdots,n_r\left(\vec{k}\right),\cdots\right\rangle = \left|n_{r_1}\left(\vec{k}_1\right)\right\rangle\left|n_{r_2}\left(\vec{k}_2\right)\right\rangle\left|n_{r_3}\left(\vec{k}_3\right)\right\rangle\cdots$ has a total of $n_{r_1}\left(\vec{k}_1\right)$ photons with wavevector $\vec{k}_1$ and polarisation r_1, $n_{r_2}\left(\vec{k}_2\right)$ photons with wavevector $\vec{k}_2$ and polarisation r_2, etc. Equations (218) generalise to:

$$\hat{a}_r\left(\vec{k}\right)\left|\cdots n_r\left(\vec{k}\right)\cdots\right\rangle = \sqrt{n_r\left(\vec{k}\right)}\left|\cdots,n_r\left(\vec{k}\right)-1,\cdots\right\rangle, \tag{244a}$$

$$\hat{a}_r^\dagger\left(\vec{k}\right)\left|\cdots n_r\left(\vec{k}\right)\cdots\right\rangle = \sqrt{n_r\left(\vec{k}\right)+1}\left|\cdots,n_r\left(\vec{k}\right)+1,\cdots\right\rangle. \tag{244b}$$

The interpretation of all of these equations is a straightforward generalisation of that given near the end of §18.2.1. Rather than considering a single harmonic oscillator, we are dealing with a superposition of independent harmonic oscillators (cf. comment following equation (197)), one for each radiation mode specified by wave-vector $\vec{k}$ and polarisation r. The operator $\hat{a}_r\left(\vec{k}\right)$ operates on the state (243) of the quantised radiation field, reducing the "occupation" number of the mode $\left(\vec{k},r\right)$ by unity, leaving all other occupation numbers unaltered; we interpret $\hat{a}_r\left(\vec{k}\right)$ as an annihilation/destruction/absorption operator which destroys one photon in the mode $\left(\vec{k},r\right)$. A similar interpretation can be made for the creation operator $\hat{a}_r^\dagger\left(\vec{k}\right)$, which creates one photon in the mode $\left(\vec{k},r\right)$.

The state of lowest energy of the radiation field is the "vacuum state", where the occupation numbers of all modes are zero. According to (240) and (241), the vacuum state has energy $\frac{1}{2}\sum_{\vec{k}}\sum_r \hbar\omega_k$, this being an infinite constant with no physical significance; we can eliminate this constant by shifting the zero of energy to coincide with that of the vacuum state; this amounts to replacing (240) by:

$$\hat{H}_{rad} = \sum_{\vec{k}}\sum_r \hbar\omega_k \hat{a}_r^\dagger\left(\vec{k}\right)\hat{a}_r\left(\vec{k}\right). \tag{245}$$

§19. The mode phase operators[58]

Having quantised the electromagnetic field by treating the electric and magnetic fields as appropriately-defined operators, we are now ready to give some application of these ideas to the problem of so-called "phase operators" in photon optics.

In view of the length of this article, we shall find it useful to break it up into a number of sub-articles.

§19.1 Overview

When one is given a complex scalar function of space and time which describes a classical electromagnetic wave, there is no difficulty in answering the question "what is the phase of the field?". This is because, at each point (**r**,t) of space-time, the phase ϕ(**r**,t) is simply the arg of the complex number U(**r**,t), i.e. :

$$U\left(\vec{r},t\right) = \left|U\left(\vec{r},t\right)\right|\exp\left(i\phi\left(\vec{r},t\right)\right). \tag{246}$$

Similar considerations apply, at least on a formal level, to the phase of a quantum-mechanical wave Ψ(**r**,t) which is a solution to the Schrödinger equation. However, the question of constructing a *phase operator* for the *quantised electromagnetic field* is one which has stimulated much debate since Dirac first considered it in 1927.

§19.2 Dirac's attempt at a phase operator

Due to the independence of the photon excitations in the fictitious box introduced just before equation (192), we shall examine electric field modes individually. Specifically, in the next few sections, we restrict attention to a *single mode* of specified wave-vector and polarization. With this in mind, we drop subscripts from the electric and magnetic field variables and treat them as scalars.

With reference to the *classical* version of (238a), a *single mode* of the Fourier expansion of the electric field would look like :

$$E = E_0\left(e^{+i\left(\vec{x}\bullet\vec{r}-\omega_k t+\phi\right)} - e^{-i\left(\vec{x}\bullet\vec{r}-\omega_k t+\phi\right)}\right). \tag{247a}$$

The phase ϕ of this single-mode field is evidently well defined.

[58] R. Loudon, *The Quantum Theory of Light* (2nd edition), 4.8.

The situation changes when we quantize the field. Dirac was the first person to investigate how one might define a phase *operator* appropriate to a single mode of the *quantized* electric field :

$$(247b)\quad \hat{E} \equiv i\sqrt{\frac{\hbar\omega_k}{2V}}\left(\hat{a}e^{i(\vec{k}\bullet\vec{x}-\omega_k t)} - \hat{a}^\dagger e^{-i(\vec{k}\bullet\vec{x}-\omega_k t)}\right)$$

which appears in equation (238a). Evidently, (247b) could be expressed in an analogous form to (247a) if, somehow, the destruction operator $\hat{a}$ could be *factorized into a product of amplitude and phase operators* :

$$(248a)\quad \hat{a} = \hat{g}\, e^{i\hat{\phi}}.$$

Now, let us *assume* that the amplitude and phase operators $\hat{g}$ and $\hat{\phi}$ are Hermitian. *If* this is correct, then the Hermitian conjugate of equation (248a) is :

$$(248b)\quad \hat{a}^\dagger = e^{-i\hat{\phi}}\hat{g}\ .$$

If we multiply (248a) and (248b) together, we get:

$$(249)\quad \begin{aligned} \hat{a}\hat{a}^\dagger &= \hat{g}\, e^{i\hat{\phi}}e^{-i\hat{\phi}}\hat{g} \\ &= \hat{g}^2 \qquad \text{(exponentials cancel)} \\ &= \hat{a}^\dagger\hat{a} + 1 \quad \text{(because } [\hat{a},\hat{a}^\dagger] = 1\text{)} \\ &= \hat{n} + 1 \qquad (\hat{a}^\dagger\hat{a} = \text{number operator}) \end{aligned}.$$

In other words, $\hat{g}^2 = \hat{n} + 1$ and so :

$$(250)\quad \hat{g} = \sqrt{\hat{n}+1}\,.$$

Hence (248a) and (248b) become :

$$(251a)\quad \hat{a} = \sqrt{\hat{n}+1}\; e^{i\hat{\phi}}\,;$$

$$(251b)\quad \hat{a}^\dagger = e^{-i\hat{\phi}}\sqrt{\hat{n}+1}\,.$$

Equivalently, we could write equations (251) in terms of the complex exponential phase operator :

$$(252a)\quad e^{i\hat{\phi}} = \frac{1}{\sqrt{\hat{n}+1}}\hat{a}\,;$$

$$(252b)\quad e^{-i\hat{\phi}} = \hat{a}^\dagger\frac{1}{\sqrt{\hat{n}+1}}\,.$$

Such is the nature of Dirac's attempt to introduce a phase operator into the theory of the quanitzed electromagnetic field.

§19.3 Objections to the Dirac phase operator

A fatal flaw in Dirac's argument was exposed by Louisell in 1963. At the heart of Louisell's objection is the demonstration that neither $\exp(i\hat{\phi})$ and $\exp(-i\hat{\phi})$ are Hermitian, and therefore cannot represent observable properties of the electromagnetic field[59]. Further, we *assumed* Hermiticity immediately after equation (248a), rendering invalid all further arguments based on this assumption. Let us justify these statements.

We begin our exposé by considering the result of applying each phase operator to the Fock state |n>:

$$(253a)\quad \begin{aligned} e^{i\hat{\phi}}|n\rangle &= \frac{1}{\sqrt{\hat{n}+1}}\hat{a}|n\rangle \ \text{(by (252a))} \\ &\quad \text{(now use (218a))} \\ &= \begin{cases} \dfrac{1}{\sqrt{\hat{n}+1}}\sqrt{n}\,|n-1\rangle & \text{if } n \neq 0 \\ 0 & \text{if } n = 0 \end{cases} \\ &= \begin{cases} \dfrac{1}{\sqrt{(n-1)+1}}\sqrt{n}\,|n-1\rangle & \text{if } n \neq 0 \\ 0 & \text{if } n = 0 \end{cases} \\ &= \begin{cases} |n-1\rangle & \text{if } n \neq 0 \\ 0 & \text{if } n = 0 \end{cases} \\ &= (1-\delta_{no})|n-1\rangle \end{aligned};$$

and :

$$(253b)\quad \begin{aligned} e^{-i\hat{\phi}}|n\rangle &= \hat{a}^\dagger\frac{1}{\sqrt{\hat{n}+1}}|n\rangle \ \text{(by (252b))} \\ &= \hat{a}^\dagger\frac{1}{\sqrt{n+1}}|n\rangle \\ &= \sqrt{n+1}\,\frac{1}{\sqrt{n+1}}|n+1\rangle \ \text{(by (218b))} \\ &= |n+1\rangle \end{aligned}.$$

Since the Fock states are orthogonal :

$$(254)\quad \langle m|n\rangle = \delta_{mn}\,,$$

we conclude that the only non-vanishing matrix elements of the exponential phase operator are :

[59] A necessary but not sufficient condition for an operator to correspond to an *observable* (in the quantum-mechanical sense of the word) is that the said operator be Hermitian. We conclude from this that *if an operator is not Hermitian it cannot be an observable*. See A.Messiah, *Quantum Mechanics (volume 1)*, North-Holland, Amsterdam (1961) p. 188.

(255a) $\left\langle n-1\left|e^{i\hat{\phi}}\right|n\right\rangle = \left\langle n-1|n-1\right\rangle = 1$,

(255b) $\left\langle n+1\left|e^{-i\hat{\phi}}\right|n\right\rangle = \left\langle n+1|n+1\right\rangle = 1$.

Now, since every Hermitian operator $\hat{O}$ has matrix elements which satisfy the condition :

(256) $\left\langle i\left|\hat{O}\right|j\right\rangle = \left\langle j\left|\hat{O}\right|i\right\rangle^*$,

we conclude from (255a) and (255b) that neither $\exp(i\hat{\phi})$ nor $\exp(-i\hat{\phi})$ are Hermitian, and so cannot be quantum-mechanical observables.

This concludes Louisell's 1963 exposé of Dirac's 1927 attempt at constructing a phase operator for the quantised electromagnetic field.

§19.4 An answer to the objection : the phase operator of Susskind and Glogower

In 1964, Susskind and Glogower found a way of meeting Louisell's objection of the previous year. They showed that the following linear combination of the Dirac phase operators :

(257a) $\hat{\cos}\,\phi = \frac{1}{2}\left(\exp\left(i\hat{\phi}\right) + \exp\left(-i\hat{\phi}\right)\right)$,

(257b) $\hat{\sin}\,\phi = \frac{1}{2i}\left(\exp\left(i\hat{\phi}\right) - \exp\left(-i\hat{\phi}\right)\right)$,

were in fact *Hermitian*. Let us prove this.

Based on the results of (255a) and (255b), it is easy to see that the only non-vanishing matrix elements of the cosine and sine phase operators are :

(258a) $\left\langle n-1\left|\hat{\cos}\,\phi\right|n\right\rangle = \left\langle n\left|\hat{\cos}\,\phi\right|n-1\right\rangle = \frac{1}{2}$;

(258b) $\left\langle n-1\left|\hat{\sin}\,\phi\right|n\right\rangle = -\left\langle n\left|\hat{\sin}\,\phi\right|n-1\right\rangle = \frac{1}{2i}$.

►*Exercise 40 : Please derive (258a) and (258b).*

With reference to (256), we conclude that the Susskind-Glogower phase operators (257a)/(257b) are both Hermitian. Let us adopt them as the operators that represent the observable properties of the quantized electromagnetic field[60].

§19.5 Number-phase uncertainty principle for the Susskind-Glogower phase operators

We are familiar with the position-momentum and energy-time uncertainty principles. In this section, we derive an *number-phase uncertainty principle* for the Susskind-Glogower phase operators.

Before launching into our derivation of the number-phase uncertainty principle, we need to recall a theorem from elementary quantum mechanics. This theorem relates the *non-commutativity of operators* to an *associated uncertainty principle* :

> **Theorem[61] :** If two operators $\hat{A}$ and $\hat{B}$ satisfy the commutation relation :
>
> (259) $[\hat{A},\hat{B}] = i\alpha$, $\quad\alpha =$ real number,
>
> then they obey the uncertainty relation :
>
> (260) $\Delta A\,\Delta B \geq \frac{1}{2}|\alpha|$,
>
> where ΔA denotes "root-mean-square deviation in the measurement of A".

Since we are after a number-phase uncertainty relation, our theorem suggests that we should calculate the commutator of the number operator and each of the phase operators. This is done in appendix B, so we just quote the result :

(261a) $[\hat{n},\hat{\cos}\,\phi] = -i\hat{\sin}\,\phi$;

(261b) $[\hat{n},\hat{\sin}\,\phi] = +i\hat{\cos}\,\phi$.

We can now use our theorem to convert (261a) and (261b) into the *number-phase uncertainty relations*:

(262a) $\Delta n\,\Delta\cos\phi \geq \frac{1}{2}\left|\left\langle\hat{\sin}\,\phi\right\rangle\right|$;

(272b) $\Delta n\,\Delta\sin\phi \geq \frac{1}{2}\left|\left\langle\hat{\cos}\,\phi\right\rangle\right|$.

►*Exercise 41 : Please prove the commutation relation :*

---

[60] I want to point out a piece of sloppiness here. A *necessary but not sufficient* condition for an operator to be a quantum-mechanical "observable" is that it be *Hermitian*. The said operator must also possess a complete set of eigen-functions. Since we have not considered the latter point, we are not really justified in promoting the Susskind-Glogower operators to the status of "observable". Regarding necessary and sufficient conditions for an operator to be an observable, see A.Messiah, *Quantum Mechanics (volume 1)*, North-Holland, Amsterdam (1961) p. 188

[61] A.Messiah, *Quantum Mechanics (volume 1)*, North-Holland, Amsterdam (1961) pp. 299-230.

(a) $[\hat{\cos}\phi, \hat{\sin}\phi] = \frac{1}{2i}\left(\hat{a}^\dagger \frac{1}{\hat{n}+1} \hat{a} - 1 \right)$

and hence show that all matrix elements of the commutator are zero except for the diagonal ground-state element :

(b) $\left\langle 0 \left| [\hat{\cos}\phi, \hat{\sin}\phi] \right| 0 \right\rangle = \frac{1}{2i}.$

Incidentally, the non-commutativity (a) of the sine and cosine phase operators has been the basis of some criticisms.

§20. Properties of the Fock states[62]

Despite being a convenient complete set of states over which to expand other (more interesting) states of the quantized electromagnetic field, the Fock states $|n\rangle$ are not of direct significance in the interpretation of experiments since electromagnetic waves generated by practical light sources do not have a precisely defined number of photons ; the uncertainty in photon number is, by definition, equal to zero for the Fock states :

(263) $\Delta n = 0$.

With reference to (258a) and (258b), the expectation values of the Susskind-Glogower phase operators are both zero for the Fock states :

(264) $\left\langle n \left| \hat{\cos}\phi \right| n \right\rangle = \left\langle n \left| \hat{\sin}\phi \right| n \right\rangle = 0.$

What is the spread in the expectation values of the phase operators? We first calculate the expectation value of the square of the cosine phase operator :

(265a)

$\left\langle n \left| \hat{\cos}^2\phi \right| n \right\rangle$

$= \frac{1}{4}\left\langle n \left| \left(e^{i\hat{\phi}} + e^{-i\hat{\phi}} \right)\left(e^{i\hat{\phi}} + e^{-i\hat{\phi}} \right) \right| n \right\rangle$ (by (257a))

$= \frac{1}{4}\left\langle n \left| \begin{matrix} \left(\frac{1}{\sqrt{\hat{n}+1}} \hat{a} + \hat{a}^\dagger \frac{1}{\sqrt{\hat{n}+1}} \right) \\ \times \left(\frac{1}{\sqrt{\hat{n}+1}} \hat{a} + \hat{a}^\dagger \frac{1}{\sqrt{\hat{n}+1}} \right) \end{matrix} \right| n \right\rangle$ (by (252))

(only terms with an equal number of creation and annihilation operators will contribute)

$= \frac{1}{4}\left\langle n \left| \frac{1}{\sqrt{\hat{n}+1}} \hat{a}\hat{a}^\dagger \frac{1}{\sqrt{\hat{n}+1}} + \hat{a}^\dagger \frac{1}{\sqrt{\hat{n}+1}} \frac{1}{\sqrt{\hat{n}+1}} \hat{a} \right| n \right\rangle$

(now use (218) and $\hat{n}|n\rangle = n|n\rangle$)

$= \frac{1}{4}\left\langle n \left| \frac{1}{\sqrt{n+1}} \sqrt{n+1} \sqrt{n+1} \frac{1}{\sqrt{n+1}} \right| n \right\rangle$

$+ \frac{1}{4}(1-\delta_{n0})\left\langle n \left| \sqrt{n} \frac{1}{\sqrt{(n-1)+1}} \frac{1}{\sqrt{(n-1)+1}} \sqrt{n} \right| n \right\rangle$

$= \begin{cases} 1/2 & \text{if } n \neq 0 \\ 1/4 & \text{if } n = 0 \end{cases}$

Similarly, a calculation of the expectation value of the square of the sine phase operator gives :

(265b) $\left\langle n \left| \hat{\sin}^2\phi \right| n \right\rangle = \begin{cases} 1/2 & \text{if } n \neq 0 \\ 1/4 & \text{if } n = 0 \end{cases}$.

Hence, if we exclude the n = 0 case,

(266a) $\Delta\cos\phi = \sqrt{\left\langle n \left| \hat{\cos}^2\phi \right| n \right\rangle - \left\langle n \left| \hat{\cos}\phi \right| n \right\rangle^2} = \frac{1}{\sqrt{2}}$.

Similarly,

(266b) $\Delta\sin\phi = \frac{1}{\sqrt{2}}$.

Equations (264) and (266) are appropriate to a phase angle that is equally likely to have any value between 0 and 2π.

§21. The coherent photon states[63]

§21.1 Introduction and overview

The coherent states were first considered by Schrödinger in 1926, who called them *states of*

[62] R. Loudon, *The Quantum Theory of Light* (2nd edition), 4.9.

[63] R. Loudon, *The Quantum Theory of Light* (2nd edition), 4.10-4.11.

minimum uncertainty product. They were first brought into prominence by Glauber[64] in the early 1960s, who coined the term *coherent states*, and they now occupy a place of especial importance in the theory of quantum optics. The nature and importance of the coherent states will become progressively clearer during the remainder of the course.

Coherent states may be defined as those states of the quantized electromagnetic field which are *eigenfunctions of the destruction operator* (see equation (267a)). We shall see that this deceptively simple definition leads to a host of important properties enjoyed by these states, some of which we outline below.

The Susskind-Glogower phase operators considered in the previous article have something to contribute to our understanding of the physical meaning of the coherent states (and conversely). Indeed, the *coherent states have an electric field that, in the limit of high field excitation, approaches that of a classical electromagnetic field with a stable amplitude and fixed phase*. In some sense, therefore, the coherent states are the closest quantum-mechanical approach of the quantized electromagnetic field to the classical electromagnetic field.

In addition to the physical properties mentioned above, the coherent states possess convenient mathematical properties which make them an extremely useful basis for expanding both quantum-mechanical fields and quantum-mechanical operators of all types.

§21.2 The coherent states : their definition and mathematical construction

There are many ways to define a coherent state. Following Glauber, let us define the coherent state $|\alpha\rangle$ to be an *eigenfunction of the destruction operator* :

(267a) $\boxed{\hat{a}|\alpha\rangle = \alpha|\alpha\rangle}$.

For future reference, let us note that the Hermitian conjugate of (267a) gives :

(267b) $\langle a|\,\hat{a}^{\dagger} = \langle a|\,\alpha^{*}$.

Now, the label "α" on our coherent state $|\alpha\rangle$ is any complex number. The coherent states therefore form a so-called *double continuum*.

It is easily verified that the following linear combination of Fock states $|n\rangle$:

$$(268)\quad |\alpha\rangle = \exp\left(-\tfrac{1}{2}|\alpha|^2\right)\sum_{n=0}^{\infty}\frac{\alpha^n}{\sqrt{n!}}|n\rangle$$

satisfies (267), and so we shall take (268) to be the Fock-state expansion of our coherent state. Note that the states defined by (268) are normalised to unity :

$$(269)\quad \langle\alpha|\alpha\rangle = 1\,.$$

►*Exercise 42 :(a) Please show that the expansion given in equation (268) is indeed a solution to (267); (b) Rather than verifying (268) to be a solution to (267), derive (268) as a consequence of (267); (c) Show that (268) satisfies the normalization relation (269).*

The coherent states are not orthogonal, as we now show :

$$\begin{aligned}
&\langle\alpha|\beta\rangle \\
&\quad= \left(e^{-\frac{1}{2}|\alpha|^2}\sum_{n=0}^{\infty}\frac{\alpha^{*n}}{\sqrt{n!}}\langle n|\right)\left(e^{-\frac{1}{2}|\beta|^2}\sum_{n'=0}^{\infty}\frac{\beta^{n'}}{\sqrt{n'!}}|n'\rangle\right) \\
&\quad\text{(rearrange terms)} \\
&\quad= e^{-\frac{1}{2}|\alpha|^2}e^{-\frac{1}{2}|\beta|^2}\sum_{n=0}^{\infty}\sum_{n'=0}^{\infty}\frac{\alpha^{*n}\beta^{n'}}{\sqrt{n!}\sqrt{n'!}}\langle n|n'\rangle \qquad (270)\\
&\quad\text{(note that } \langle n|n'\rangle = \delta_{nn'}) \\
&\quad= \exp\left(-\tfrac{1}{2}|\alpha|^2 - \tfrac{1}{2}|\beta|^2\right)\sum_{n=0}^{\infty}\frac{(\alpha^{*}\beta)^n}{n!} \\
&\quad\text{(use Taylor expansion of exp(x))} \\
&\quad= \exp\left(-\tfrac{1}{2}|\alpha|^2 - \tfrac{1}{2}|\beta|^2 + \alpha^{*}\beta\right)
\end{aligned}$$

Despite their lack of orthogonality, the coherent states are *complete*. In fact, they are *overcomplete*, and their lack of orthogonality is a consequence of this.

[64] R.J.Glauber, *The quantum theory of optical coherence*, Phys. Rev. **130** 2529-2539 (1963) ; R.J.Glauber, *Coherent and incoherent states of the radiation field*, Phys. Rev. **131** 2766-2788 (1963).

§21.3 Properties of the coherent states

In this section we explore some of the fundamental properties of the coherent states. We shall look at :

(a) The mean & uncertainty in photon number ;

(b) The probability of measuring n photons ;

(c) The mean & uncertainty of the S-G phase ;

(d) The mean & uncertainty of the electric field.

§21.3.1 Mean & uncertainty of photon number

Let us calculate the mean number of photons in a given coherent state $|\alpha\rangle$:

$$\begin{aligned}\langle\alpha|\hat{n}|\alpha\rangle &= \langle\alpha|\hat{a}^\dagger\hat{a}|\alpha\rangle \\ &\quad \text{(now use (267a)/(267b))} \\ &= \langle\alpha|\alpha^*\alpha|\alpha\rangle \\ &= |\alpha|^2\langle\alpha|\alpha\rangle \\ &= |\alpha|^2\end{aligned} \tag{271}$$

With a view to obtaining the root-mean-square deviation Δn in the photon number, let us now calculate the expectation value of the square of the number operator :

$$\begin{aligned}\langle\alpha|\hat{n}^2|\alpha\rangle &= \langle\alpha|\hat{a}^\dagger\hat{a}\hat{a}^\dagger\hat{a}|\alpha\rangle \\ &\quad \text{(now use (267a)/(267b))} \\ &= \langle\alpha|\alpha^*\hat{a}\hat{a}^\dagger\alpha|\alpha\rangle \\ &\quad \text{(now use (203))} \\ &= \alpha^*\alpha\langle\alpha|(\hat{a}^\dagger\hat{a}+1)|\alpha\rangle \\ &\quad \text{(use (267a)/(267b) again)} \\ &= \alpha^*\alpha\langle\alpha|(\alpha^*\alpha+1)|\alpha\rangle \\ &= \alpha^*\alpha(\alpha^*\alpha+1)\langle\alpha|\alpha\rangle \\ &= |\alpha|^2+|\alpha|^4\end{aligned} \tag{272}$$

We conclude from (271) and (272) that the root-mean-square deviation in the number of photons n in a coherent state is nonzero :

$$\begin{aligned}\Delta n &= \sqrt{\langle\alpha|\hat{n}^2|\alpha\rangle-\langle\alpha|\hat{n}|\alpha\rangle^2} \\ &= \sqrt{|\alpha|^2+|\alpha|^4-|\alpha|^4} \\ &= |\alpha|\end{aligned} \tag{273}$$

Combining (271) and (273), we see that the *fractional uncertainty* in the number of photons in the coherent cavity mode is :

$$\frac{\Delta n}{\langle\alpha|\hat{n}|\alpha\rangle} = \frac{|\alpha|}{|\alpha|^2} = \frac{1}{|\alpha|} . \tag{274}$$

This tends to zero with increasing excitation of the coherent cavity mode.

§21.3.2 Probability of measuring n photons

Let us calculate the probability of finding n photons in the coherent cavity mode :

$$\begin{aligned}\left|\langle n|\alpha\rangle\right|^2 &= \left|\langle n|\exp\left(-\tfrac{1}{2}|\alpha|^2\right)\sum_{n'=0}^{\infty}\frac{\alpha^{n'}}{\sqrt{n'!}}|n'\rangle\right|^2 \\ &= \left|\exp\left(-\tfrac{1}{2}|\alpha|^2\right)\sum_{n'=0}^{\infty}\frac{\alpha^{n'}}{\sqrt{n'!}}\langle n|n'\rangle\right|^2 \\ &\quad \text{(only one term survives)} \\ &= \left|\exp\left(-\tfrac{1}{2}|\alpha|^2\right)\frac{\alpha^n}{\sqrt{n!}}\right|^2 \\ &= \exp\left(-|\alpha|^2\right)\frac{|\alpha|^{2n}}{n!}\end{aligned} \tag{275}$$

This is a *Poisson distribution* with mean and spread consistent with equations (271) and (273).

§21.3.3 Mean & uncertainty of the Susskind-Glogower phase

The expectation values and uncertainties of the Susskind-Glogower phase operators are unfortunately rather complicated, and so we restrict attention to cosine phase operator.

Before commencing our calculations, let us define θ to be the phase of the complex number α which labels the coherent state :

$$\alpha = |\alpha|e^{i\theta} . \tag{276}$$

With the help of (255a), (255b), (268) and (276), one can show that :

$$\begin{aligned}\langle\alpha|\hat{c}\cos\phi|\alpha\rangle &= \frac{e^{-|\alpha|^2}}{2}\sum_{n=0}^{\infty}\frac{(\alpha^*)^{n+1}\alpha^n+(\alpha^*)^n\alpha^{n+1}}{\sqrt{n!(n+1)!}} \\ &= |\alpha|e^{-|\alpha|^2}\cos\theta\sum_{n=0}^{\infty}\frac{|\alpha|^{2n}}{n!\sqrt{n+1}} .\end{aligned} \tag{277}$$

►*Exercise 43 : Please prove (277).*

We can see from (277) that the expectation value of $\hat{\cos}\phi$ is proportional to the phase of the complex number α which labels a given coherent state. One can also show that :

$$\left\langle \alpha \left| \hat{\cos}^2 \phi \right| \alpha \right\rangle = \tfrac{1}{2} - \tfrac{1}{4} e^{-|\alpha|^2} + |\alpha|^2 \left(\cos^2\theta - \tfrac{1}{2} \right) e^{-|\alpha|^2} \sum_{n=0}^{\infty} \frac{|\alpha|^{2n}}{n!\sqrt{(n+1)(n+2)}} . \quad (278)$$

►*Exercise 44: Please prove (278).*

Unfortunately, nobody knows how to explicitly evaluate the sums which appear in (277) and (278). Since we are interested in the phase-behavior of the coherent states when | α |>>1, we can make use of the following asymptotic expansions :

$$\sum_{n=0}^{\infty} \frac{|\alpha|^{2n}}{n!\sqrt{n+1}} \sim \frac{\exp\left(|\alpha|^2\right)}{|\alpha|} \left(1 - \frac{1}{8|\alpha|^2} + \ldots \right), \quad |\alpha| >> 1 \quad ; \quad (279a)$$

$$\sum_{n=0}^{\infty} \frac{|\alpha|^{2n}}{n!\sqrt{(n+1)(n+2)}} \sim \frac{\exp\left(|\alpha|^2\right)}{|\alpha|^2} \left(1 - \frac{1}{2|\alpha|^2} - \ldots \right), \quad |\alpha| >> 1. \quad (279b)$$

With the help of these expansions, let us write down approximate forms for equation (277) :

$$\left\langle \alpha \left| \hat{\cos}\phi \right| \alpha \right\rangle \approx |\alpha| e^{-|\alpha|^2} \cos\theta \frac{\exp\left(|\alpha|^2\right)}{|\alpha|} \left(1 - \frac{1}{8|\alpha|^2} + \ldots \right) = \cos\theta \left(1 - \frac{1}{8|\alpha|^2} + \ldots \right), \quad |\alpha >> 1. \quad (280)$$

and for equation (278) :

$$\left\langle \alpha \left| \hat{\cos}^2 \phi \right| \alpha \right\rangle \approx \tfrac{1}{2} - \tfrac{1}{4} e^{-|\alpha|^2} + |\alpha|^2 \left(\cos^2\theta - \tfrac{1}{2} \right) e^{-|\alpha|^2} \frac{e^{+|\alpha|^2}}{|\alpha|^2} \left(1 - \frac{1}{2|\alpha|^2} - \ldots \right)$$

$$= \tfrac{1}{2} - \tfrac{1}{4} e^{-|\alpha|^2} + \left(\cos^2\theta - \tfrac{1}{2} \right) \left(1 - \frac{1}{2|\alpha|^2} - \ldots \right) \quad (281)$$

$$= -\tfrac{1}{4} e^{-|\alpha|^2} + \cos^2\theta - \frac{1}{2|\alpha|^2} \left(\cos^2\theta - \tfrac{1}{2} \right) - \ldots$$

$$\approx \cos^2\theta - \frac{\left(\cos^2\theta - \frac{1}{2} \right)}{2|\alpha|^2} - \ldots, \quad |\alpha| >> 1$$

We are finally able to calculate the uncertainty in the cosine phase for the coherent states :

$$\Delta\cos\phi = \sqrt{\left\langle \alpha \left| \hat{\cos}^2 \phi \right| \alpha \right\rangle - \left\langle \alpha \left| \hat{\cos}\phi \right| \alpha \right\rangle^2} \quad (282)$$

(now use (280) and (281))

$$= \sqrt{\cos^2\theta - \frac{\left(\cos^2\theta - \frac{1}{2} \right)}{2|\alpha|^2} - \cos^2\theta \left(1 - \frac{1}{8|\alpha|^2} + \ldots \right)^2}$$

$$= \sqrt{\cos^2\theta - \frac{\left(\cos^2\theta - \frac{1}{2} \right)}{2|\alpha|^2} - \cos^2\theta + \frac{\cos^2\theta}{4|\alpha|^2} - \ldots}$$

$$= \sqrt{\frac{1 - \cos^2\theta}{4|\alpha|^2} - \ldots} = \sqrt{\frac{\sin^2\theta}{4|\alpha|^2} - \ldots} \approx \frac{|\sin\theta|}{2|\alpha|}.$$

Evidently, the uncertainty in the Susskind-Glogower cosine phase operator becomes *zero* as $|\alpha| \to \infty$. Also, (273) and (282) lead to the following expression for the uncertainty product :

$$\Delta n \Delta\cos\phi = \tfrac{1}{2} |\sin\theta|, \quad |\alpha| >> 1 | . \quad (283)$$

On comparing (283) to the amplitude-phase uncertainty relation (262), we conclude that the coherent states achieves the theoretical *minimum uncertainty product* for number and phase. This is yet another reason for these states being termed "coherent".

Recalling (274) and (282) we see that, as the mean photon number is increased, the coherent states become progressively better defined in terms of both amplitude and phase angle.

§21.3.4 Mean & uncertainty of electric field

With reference to equation (238a), let us calculate the expectation value of the single-mode electric field operator for a coherent state :

(284)

$$\begin{aligned}\left\langle\alpha\middle|\hat{E}\middle|\alpha\right\rangle &= i\sqrt{\frac{\hbar\omega}{2V}}\left\langle\alpha\middle|\hat{a}e^{i\left(\vec{k}\bullet\vec{x}-\omega t\right)}-\hat{a}^{\dagger}e^{-i\left(\vec{k}\bullet\vec{x}-\omega t\right)}\middle|\alpha\right\rangle\\ &= i\sqrt{\frac{\hbar\omega}{2V}}\left\langle\alpha\middle|\alpha e^{i\left(\vec{k}\bullet\vec{x}-\omega t\right)}-\alpha^{*}e^{-i\left(\vec{k}\bullet\vec{x}-\omega t\right)}\middle|\alpha\right\rangle\\ &= i\sqrt{\frac{\hbar\omega}{2V}}\left(\alpha e^{i\left(\vec{k}\bullet\vec{x}-\omega t\right)}-\alpha^{*}e^{-i\left(\vec{k}\bullet\vec{x}-\omega t\right)}\right)\left\langle\alpha\middle|\alpha\right\rangle\\ &= i\sqrt{\frac{\hbar\omega}{2V}}\left(|\alpha|e^{i\left(\vec{k}\bullet\vec{x}-\omega t+\theta\right)}-|\alpha|e^{-i\left(\vec{k}\bullet\vec{x}-\omega t+\theta\right)}\right)\\ &= -2\sqrt{\frac{\hbar\omega}{2V}}|\alpha|\sin\left(\vec{k}\bullet\vec{x}-\omega t+\theta\right)\end{aligned}$$

With reference to (247a) we see that, up to constants of proportionality, *this expression is identical to that of a classical single-mode electromagnetic field disturbance.* The polar angle θ of the complex number α is evidently the quantum mechanical analogue of the classical phase factor ϕ which appears in (247a).

Let us now turn to the root-mean-square fluctuations in the expectation value of the electric field. Using similar arguments to those given above, we may calculate the expectation value of the square of the electric field:

$$(285)\quad \left\langle\alpha\middle|\hat{E}^{2}\middle|\alpha\right\rangle = \frac{\hbar\omega}{2V}\left(4|\alpha|^{2}\sin^{2}\left(\vec{k}\bullet\vec{r}-\omega t+\theta\right)+1\right).$$

►*Exercise 45: Please prove equation (285).*

Hence, in the usual manner, we may calculate the root-mean-square deviation in the electric field :

$$\begin{aligned}\Delta E &= \sqrt{\left\langle\alpha\middle|\hat{E}^{2}\middle|\alpha\right\rangle-\left\langle\alpha\middle|\hat{E}\middle|\alpha\right\rangle^{2}}\\ (286)\qquad &= \sqrt{\begin{matrix}\frac{\hbar\omega}{2V}\left(4|\alpha|^{2}\sin^{2}\left(\vec{k}\bullet\vec{x}-\omega t+\theta\right)+1\right)\\ -4\frac{\hbar\omega}{2V}|\alpha|^{2}\sin^{2}\left(\vec{k}\bullet\vec{x}-\omega t+\theta\right)\end{matrix}}.\\ &= \sqrt{\frac{\hbar\omega}{2V}}\end{aligned}$$

While (284) indicates that the expectation value of the electric field increases proportionally to $|\alpha|$, we see from (286) that the *root-mean-square deviation in the electric field is independent of the level of excitation of the coherent state*. Thus, not only does the expectation value of the electric field for a coherent state look exactly like the disturbance due to a classical single-mode field, but the *relative* uncertainty in the electric field becomes zero as the degree of excitation of the coherent state becomes very large. We conclude that the coherent states make a smooth transition from the quantum to the classical domains. This may be contrasted to the case of the Fock states which was studied in an earlier article, which made no such smooth transition from the quantum to the classical domains.

►*Exercise 46: Using the formulae given in lectures, show that the coherent states* $|\alpha\rangle$ *may be written in terms of a certain operator acting on the Fock vacuum* $|0\rangle$ *:*

$$(287)\quad |\alpha\rangle = \exp\left(-\tfrac{1}{2}|\alpha|^{2}\right)\sum_{n=0}^{\infty}\frac{\left(\alpha\hat{a}^{\dagger}\right)^{n}}{n!}|0\rangle.$$

Hence show that :

$$(288)\quad |\alpha\rangle = \exp\left(\alpha\hat{a}^{\dagger}-\tfrac{1}{2}|\alpha|^{2}\right)|0\rangle.$$

It can be shown that, for two operators $\hat{c}$ *and* $\hat{d}$ *, which commute with their commutator :*

$$(289)\quad [\hat{c},[\hat{c},\hat{d}]]=[\hat{d},[\hat{c},\hat{d}]]=0,$$

the following relation holds :

$$(290)\quad \exp(\hat{c})\exp(\hat{d}) = \exp\left(\hat{c}+\hat{d}+\tfrac{1}{2}[\hat{c},\hat{d}]\right).$$

Bearing this in mind, please show that (288) can be written as :

$$(291)\quad |\alpha\rangle = \exp\left(\alpha\hat{a}^{\dagger}-\alpha^{*}\hat{a}\right)|0\rangle.$$

► *Exercise 47: Despite their non-orthogonality, coherent states are complete (in the mathematical sense of the word). To see this, prove the completeness relation for coherent states :*

$$(292)\quad \frac{1}{\pi}\iint|\alpha\rangle\langle\alpha|\,d^{2}\alpha = 1,$$

where the integral is taken over the entire complex plane. Hints : Use the Fock expansion of the coherent states & transform to polar coordinates. You may need the integral :

(293) $2\int_0^\infty e^{-r^2}\, r^{2n+1} dr = n!$

Show that an arbitrary operator $\hat{O}$ possesses the following integral representation in terms of the coherent states :

(294) $\hat{O} = \frac{1}{\pi^2}\iiint\!\!\iint \langle\alpha'|\hat{O}|\alpha''\rangle |\alpha'\rangle\langle\alpha''|\, d^2\alpha'\, d^2\alpha''$.

►*Exercise 48: Consider the single-mode state $|\theta\rangle$ which is defined by the following expansion over the Fock states $|n\rangle$:*

(295) $|\theta\rangle = \frac{1}{\sqrt{s+1}}\sum_{n=0}^{s} e^{in\theta}|n\rangle$,

where s is an integer and θ is a real number.
(a) Please calculate the expectation value in the photon number. You may need to make use of the following result for summation of an arithmetic series:

(296) $\sum_{m=0}^{s} m = \frac{1}{2}s(s+1)$.

(b) Next, please determine the expectation value of the Susskind-Glogower cosine phase operator for the states $|\theta\rangle$. You may find relations (253a) and (253b) from the lectures to be useful.
(c) Lastly, please give a brief interpretation of the result you obtained in (b).

►*Exercise 49: In lectures, we saw that a number of useful and interesting properties followed from the definition of the coherent states as the eigenfunctions of the destruction operator. As a variation on this theme, consider the intelligent states $|\alpha,\beta,\xi\rangle$ of a single-mode field, which are by definition equal to the eigenfunctions of the operator $\alpha\hat{a}^2 + \beta\hat{a}^{\dagger 2}$:*

(297) $\left(\alpha\hat{a}^2 + \beta\hat{a}^{\dagger 2}\right)|\alpha,\beta,\xi\rangle = \xi|\alpha,\beta,\xi\rangle$,

where α and β are real numbers, and ξ is a complex number. Suppose the intelligent states to possess the following expansion over the Fock states $|2n\rangle$ of even occupation number:

(298) $|\alpha,\beta,\xi\rangle = \sum_{n=0}^{\infty} C_{2n}|2n\rangle$.

By substituting this Fock expansion into the defining relation for the intelligent states, please show that the coefficients C_{2n} of this expansion satisfy the recursion relation:

(299) $-\xi C_n + \alpha\sqrt{(n+1)(n+2)}\,C_{n+2} + \beta\sqrt{n(n-1)}\,C_{n-2} = 0$

►*Exercise 50: The so-called "Garden" states $|\xi\rangle$ are defined by:*

(300) $\left(\hat{a}\cosh\theta + \hat{a}^\dagger \sinh\theta\right)|\xi\rangle = \xi|\xi\rangle$,

where $\hat{a}$ and $\hat{a}^\dagger$ are the annihilation and creation operators respectively, θ is a real number, and ξ is a complex number. Consider the expansion of the Garden states $|\xi\rangle$ over the Fock states $|n\rangle$:

(301) $|\xi\rangle = \sum_{n=0}^{\infty}\xi_n|n\rangle$.

Please determine the recursion relation satisfied by the expansion coefficients ξ_n.

APPENDIX B

In this appendix, we derive equations (261a) and (261b), namely :

(261a) $[\hat{n}, \hat{\cos}\phi] = -i\hat{\sin}\phi$;

(261b) $[\hat{n}, \hat{\sin}\phi] = +i\hat{\cos}\phi$.

We begin the derivation by calculating the value of four different commutators :

First commutator :

$$\begin{aligned}
&\left[\hat{n},\hat{a}^\dagger\right] = \hat{n}\hat{a}^\dagger - \hat{a}^\dagger\hat{n} \\
&\quad = \hat{a}^\dagger\hat{a}\hat{a}^\dagger - \hat{a}^\dagger\hat{a}^\dagger\hat{a} \ \text{(because } \hat{n} = \hat{a}^\dagger\hat{a}) \\
&\quad = \hat{a}^\dagger\left(\hat{a}\hat{a}^\dagger - \hat{a}^\dagger\hat{a}\right) \ \text{(factor out } \hat{a}^\dagger \text{ on left)} \\
&\quad = \hat{a}^\dagger\left[\hat{a},\hat{a}^\dagger\right] \\
&\quad = \hat{a}^\dagger \ (\text{because } \left[\hat{a},\hat{a}^\dagger\right] = 1)
\end{aligned} \tag{A1}$$

Second commutator :

(A2)
$$\begin{aligned}[\hat{n},\hat{a}] &= \hat{n}\hat{a}-\hat{a}\hat{n}\\ &= \hat{a}^\dagger\hat{a}\hat{a}-\hat{a}\hat{a}^\dagger\hat{a} \text{ (because } \hat{n}=\hat{a}^\dagger\hat{a})\\ &= \left(\hat{a}^\dagger\hat{a}-\hat{a}\hat{a}^\dagger\right)\hat{a} \text{ (factor out } \hat{a} \text{ on right)}\\ &= \left[\hat{a}^\dagger,\hat{a}\right]\hat{a}\\ &= -\left[\hat{a},\hat{a}^\dagger\right]\hat{a}\\ &= -\hat{a} \text{ (because } \left[\hat{a},\hat{a}^\dagger\right]=1)\end{aligned}$$

Third commutator :

(A3)
$$\left[\hat{n},e^{i\hat{\phi}}\right]=\left[\hat{n},\frac{1}{\sqrt{\hat{n}+1}}\hat{a}\right] \text{ (using equation (252a))}$$
$$=\left[\hat{n},\frac{1}{\sqrt{\hat{n}+1}}\right]\hat{a}+\frac{1}{\sqrt{\hat{n}+1}}\left[\hat{n},\hat{a}\right]$$
("[A,BC] = [A,B]C + B[A,C]" has been used above)
$$=0+\frac{1}{\sqrt{\hat{n}+1}}\left[\hat{n},\hat{a}\right]$$
$$=\frac{-1}{\sqrt{\hat{n}+1}}\hat{a} \text{ (by (A2))}$$
$$=-e^{i\hat{\phi}} \text{ (using equation (252a) again)}$$

Fourth commutator :

(A4)
$$\left[\hat{n},e^{-i\hat{\phi}}\right]=\left[\hat{n},\hat{a}^\dagger\frac{1}{\sqrt{\hat{n}+1}}\right] \text{ (using equation (252b))}$$
$$=\left[\hat{n},\hat{a}^\dagger\right]\frac{1}{\sqrt{\hat{n}+1}}+\hat{a}^\dagger\left[\hat{n},\frac{1}{\sqrt{\hat{n}+1}}\right]$$
("[A,BC] = [A,B]C + B[A,C]" has been used above)
$$=\left[\hat{n},\hat{a}^\dagger\right]\frac{1}{\sqrt{\hat{n}+1}}+0$$
$$=\hat{a}^\dagger\frac{1}{\sqrt{\hat{n}+1}} \text{ (by (A1))}$$
$$=e^{-i\hat{\phi}} \text{ (using equation (252b) again)}$$

We are now ready to do our proof.

Proof of (261a) :
$$[\hat{n},\hat{\cos}\phi]=\tfrac{1}{2}[\hat{n},e^{i\hat{\phi}}+e^{-i\hat{\phi}}]$$
$$=\tfrac{1}{2}[\hat{n},e^{i\hat{\phi}}]+\tfrac{1}{2}[\hat{n},e^{-i\hat{\phi}}]$$
$$=-\tfrac{1}{2}e^{i\hat{\phi}}+\tfrac{1}{2}e^{-i\hat{\phi}} \text{ (using (A3) and (A4))}$$
$$=-i\tfrac{1}{2i}\left(e^{i\hat{\phi}}-e^{-i\hat{\phi}}\right)$$
$$=-i\hat{\sin}\phi \quad Q.E.D.$$

Proof of (261b):

(A6)
$$[\hat{n},\hat{\sin}\phi]=\tfrac{1}{2i}[\hat{n},e^{i\hat{\phi}}-e^{-i\hat{\phi}}]$$
$$=\tfrac{1}{2i}[\hat{n},e^{i\hat{\phi}}]-\tfrac{1}{2i}[\hat{n},e^{-i\hat{\phi}}]$$
$$=-\tfrac{1}{2i}e^{i\hat{\phi}}-\tfrac{1}{2i}e^{-i\hat{\phi}} \text{ (using (A3) and (A4))} .$$
$$=\tfrac{i}{2}\left(e^{i\hat{\phi}}+e^{-i\hat{\phi}}\right)$$
$$=+i\hat{\cos}\phi \quad Q.E.D.$$

CHAPTER 4
QUANTUM THEORY OF COHERENCE

§22. Introduction and overview

In this section of the course we shall study Glauber's quantum theory of coherence[65] and apply it to Young's double-slit experiment from the point of view of the theory of the quantized electromagnetic field. We then introduce the quantum correlation functions of arbitrary order and hence show that the coherent states are coherent to all orders.

There are a number of reasons for including this material in the course. (i) In classical statistical mechanics, probabilities enter into a strictly deterministic theory, because one has incomplete knowledge about a given system of interest. For example, one may not know (or wish to know) the position and momenta of each atom in a glass of water, when studying such a system from the point of view of classical statistical mechanics. In quantum statistical mechanics, probabilities enter at two different levels. Firstly, one has the intrinsic indeterminacy of quantum systems, whereby a (completely known) wavefunction yields the probabilities associated with certain observables. Secondly, for a quantum system whose wavefunction is only partially known, there is a second level of indeterminacy. The density-matrix formalism, which will be introduced in the next section, gives a means for dealing with this dual uncertainty, in the context of quantum statistical mechanics. This formalism finds widespread use

[65] R.J.Glauber, *The quantum theory of optical coherence*, Phys.Rev. **130** 2529-2539 (1963) ; R.J.Glauber, *Coherent and incoherent states of the radiation field*, Phys.Rev. **131** 2766-2788 (1963).

throughout physics, examples of which include the theory of Bose-Einstein condensation, superfluidity, quantum optics, atom optics, strongly correlated electron systems at finite temperature, superconductivity, quantum theory of solids *etc.*
(ii) Our previous studies, of both the coherent states and the multi-mode quantised electromagnetic field, will be further deepened in the present material.
(iii) The area of quantum optics, which is receiving considerable impetus as a supporting technology for quantum computation, in addition to being a vibrant field in its own right, has at its very heart the theory of quantum coherence. Indeed, the work of Glauber, whose essence will be studied in the ensuing paragraphs, is often cited as ushering in the field of quantum optics. Accordingly, I felt that a quantum mechanics course should include a small amount of quantum optics, giving you the basic language and concepts needed to read further in this subject, if you desire.

§23. The density operator[66]

We are all familiar with the distinction between classical and statistical mechanics. A similar distinction exists between the quantum mechanics of the so-called *pure states* and the quantum mechanics of *statistical mixtures*.

Pure states: The precisely-specified wavefunctions we have been dealing with, such as the Fock and coherent states, together with the Dirac and Klein-Gordon wavefunctions, are examples of pure states. In some sense, knowledge of the whole wavefunction corresponds to us having maximal knowledge of the quantum state. Intrinsic quantum uncertainties of course exist, but this does not contradict the previous statement regarding maximal knowledge.

Statistical mixtures: In addition to intrinsic quantum uncertainty, one may simply be in ignorance of the initial state of a quantum system. Specifically, rather than having a single wavefunction (pure state), one may have an *ensemble* of wavefunctions, each of which have an associated probability of being the correct wavefunction. This is the essential nature of the quantum-mechanical statistical mixture, which may be described by its *density matrix*. One would still like to be able to perform calculations on quantum systems of which one has only partial knowledge, and the formalism of the density matrix provides a means to do this.

An example of a statistical mixture is the thermal excitation of a cavity mode which is studied while deriving the famous Planck radiation law[67]. There, we see that the probability P_n that n photons occupy the energy level En at temperature T is[68]:

$$(302)\quad P_n = \frac{\exp\left(E_n / (k_B T)\right)}{\sum_n \exp\left(E_n / (k_B T)\right)} .$$

In a pure state |R>, the expectation value for an observable represented by the operator $\hat{O}$ is:

$$(303)\quad \left\langle \hat{O} \right\rangle = \left\langle R \middle| \hat{O} \middle| R \right\rangle .$$

For the statistical mixture, we have instead a weighted sum over our ensemble:

$$(304)\quad \left\langle \hat{O} \right\rangle = \sum_R P_R \left\langle R \middle| \hat{O} \middle| R \right\rangle$$

where the probabilities P_R are normalized to unity:

$$(305)\quad \sum_R P_R = 1 .$$

Now, in at least one of your quantum mechanics courses many of you would have learnt of the so-called *closure relation* which constitutes a statement of the *completeness* (in the mathematical sense of the word) of a set of state vectors {|S>}[69]:

$$(306)\quad \sum_S |S\rangle\langle S| = 1 .$$

If we insert this relation immediately after the $\hat{O}$ in (304), we arrive at:

$$(307)\quad \begin{aligned} \left\langle \hat{O} \right\rangle &= \sum_R P_R \langle R | \hat{O} | \sum_S |S\rangle\langle S| R \rangle \\ &= \sum_S \langle S | \left\{ \sum_R P_R |R\rangle\langle R| \right\} \hat{O} | S \rangle \end{aligned} .$$

We now define the so-called *density operator*[70] by:

[66] See Messiah, *Quantum Mechanics volume 1*, pp. 204-5 and 331-338.

[67] More generally, the wavefunction for many finite-temperature condensed matter or radiation systems, can be described in terms of its associated density matrix.

[68] We will return to such a construct, for those of you who will be taking the Advanced Quantum Mechanics course.

[69] Ask me if you would like a simple geometric interpretation for this expression. Also, I may give you such an interpretation, even if you do not ask for it!

[70] Note that the existence of non-zero off-diagonal matrix elements, of the density matrix, is very closely related to the concept of "off diagonal long range order", which forms an important foundational concept in many theories of condensed

(308) $\hat{\rho} = \sum_R P_R |R\rangle\langle R|$

and so (307) becomes:

(309) $$\langle \hat{O} \rangle = \sum_S \langle S | \hat{\rho}\hat{O} | S \rangle = Trace(\hat{\rho}\hat{O})$$

where the trace of an operator is defined to be the sum of its diagonal matrix elements in any representation.

The formalism leading to (309) has the advantage of the simplicity in the manipulations required to produce expectation values when one is dealing with statistical mixtures. It is completely equivalent to the idea of an ensemble of wave-functions, which was mentioned earlier in this section.

Remark 1: Setting $\hat{O}=1$ in (309), we learn that the density matrix has trace unity:

(310) $\text{Trace}(\hat{\rho}) = 1$.

Remark 2: For pure states, $P_R = 1$ for one and only one value of R, and so (308) reduces to:

(311) $\hat{\rho} = |R\rangle\langle R|$.

Remark 3: A statistically-stationary[71] quantum-mechanical field is one for which the density matrix does not change with time:

(312) $\partial_t \hat{\rho} = 0$.

An equivalent condition to (312) is that the density operator commute with the Hamiltonian:

(313) $[H, \hat{\rho}] = 0$.

matter physics. Such theories include the theory of superfluidity, and the theory of Bose-Einstein condensation. For an example of the latter application of the concept of off-diagonal long-range order, see Piatevskii and Stringari, *Bose-Einstein Condensation*, Oxford University Press, Oxford (2003). I would be happy to lend you a copy of this book, if you are interested in pursuing this point further. I also have a reference for off-diagonal long range order in the context of the theories of superconductivity and superfluidity, which I am also happy to lend to anyone who is interested!

[71] By "statistically stationary", we mean that the statistics of the system do not vary with time. I will probably give a couple of examples of such systems, during the course of the lecture. If I forget to do this, please ask me!

§24. The intensity operator for the quantized electromagnetic field[72]

To date, our discussions of the quantized electromagnetic field have considered the case of the *free* field, *i.e.*, the electromagnetic field in a volume of space which is free of both currents and charges. However, experiments concerned with the detection of radiation involve the interaction of the quantized electromagnetic field with atoms in the detector. For example, the detection of a photon evidently involves the *absorption* of the said photon from the electromagnetic field under study. Thus we must go beyond our free-field theory in order to develop a (crude!) quantum theory of detection. Here, we develop a simple form of such a theory, in which the so-called *intensity operator* plays a central role[73].

Before proceeding with the main topic of this article, it shall prove convenient to divide the electric-field operator given in (238a) into two parts:

(314) $\hat{\vec{E}}(\vec{x},t) = \hat{\vec{E}}^{+}(\vec{x},t) + \hat{\vec{E}}^{-}(\vec{x},t)$,

where:

(315a) $\hat{\vec{E}}^{+}(\vec{x},t) \equiv \sum_{\vec{k}} i\sqrt{\frac{\hbar\omega_k}{2V}}\vec{\varepsilon}_r(\vec{k})\hat{a}_r(\vec{k})e^{i(\vec{k}\bullet\vec{x}-\omega_k t)}$,

(315b) $\hat{\vec{E}}^{-}(\vec{x},t) \equiv -\sum_{\vec{k}} i\sqrt{\frac{\hbar\omega_k}{2V}}\vec{\varepsilon}_r(\vec{k})\hat{a}_r^{\dagger}(\vec{k})e^{-i(\vec{k}\bullet\vec{x}-\omega_k t)}$.

As we mentioned earlier, the process of detection involves photon *absorption*; when making an intensity measurement on the electromagnetic field, the initial and final states of the said field will therefore differ. If the electromagnetic field makes a transition from an initial state |i> to a final state |f> in which one (detected) photon of wave-vector $\vec{k}$ has been removed, the matrix element associated with the transition[74] takes the form:

[72] The argument given here is rather heuristic and follows that given in Glauber's paper on quantum coherence which was cited earlier. These heuristic arguments are repeated in many texts on quantum optics: see, for example, pp. 573-575 of L.Mandel and E.Wolf, *Optical Coherence and Quantum Optics*, Cambridge University Press (1995).

[73] For a much more detailed study of the interaction of the quantized electromagnetic field with matter, see the first section of the Advanced Quantum Mechanics course.

[74] The relevance of this matrix element will become clearer if you briefly revise the theory of time-dependent perturbation

(316) $\left\langle f\left|\hat{\vec{E}}^{+}(\vec{x},t)\right|i\right\rangle$.

An important point to note is that we have only included the absorption component of the electric-field operator: this is because *real transitions are only possible in which a photon has been removed from the field* [75]. The count-rate R which would be registered on an ideal photon detector of negligible size and frequency-independent photo-absorption probability is proportional to the modulus squared of this quantity[76]:

(317) $R(\vec{x},t) = D\sum_f \left|\left\langle f\left|\hat{\vec{E}}^{+}(\vec{x},t)\right|i\right\rangle\right|^2$,

where D is a detector-dependent constant whose functional form does not concern us. We have summed over all possible final states |f> because this final state of the field usually remains unobserved[77].

Let us express (317) in a more convenient form:

$$R(\vec{x},t) = D\sum_f \left\langle f\left|\hat{\vec{E}}^{+}(\vec{x},t)\right|i\right\rangle\left\langle i\left|\hat{\vec{E}}^{-}(\vec{x},t)\right|f\right\rangle$$

(swap terms in sum around)

$$= D\sum_f \left\langle i\left|\hat{\vec{E}}^{-}(\vec{x},t)\right|f\right\rangle\left\langle f\left|\hat{\vec{E}}^{+}(\vec{x},t)\right|i\right\rangle$$

(318) (now reorder)

$$= D\left\langle i\left|\hat{\vec{E}}^{-}(\vec{x},t)\right\{\sum_f |f\rangle\langle f|\right\}\hat{\vec{E}}^{+}(\vec{x},t)|i\rangle$$

(now use the closure relation (306))

$$= D\left\langle i\left|\hat{\vec{E}}^{-}(\vec{x},t)\hat{\vec{E}}^{+}(\vec{x},t)\right|i\right\rangle.$$

Since count-rate is proportional to intensity, we may therefore write down the *intensity operator* associated with the quantized electromagnetic field:

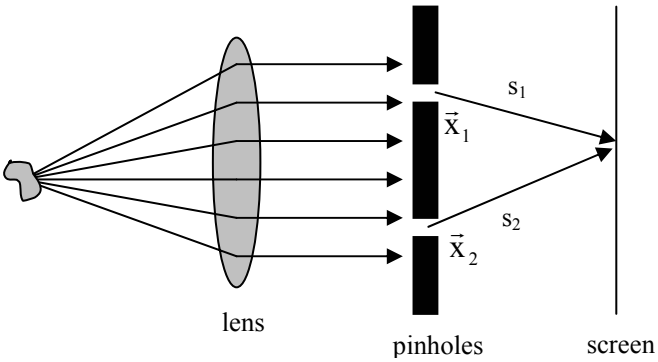
(319) $\hat{I}(\vec{x},t) = J\hat{\vec{E}}^{-}(\vec{x},t)\hat{\vec{E}}^{+}(\vec{x},t)$,

where J is a positive constant whose form does not concern us here. Result (319) is valid for a pure state. When the light beam is in a mixed state (e.g. if we have a statistical source) described by a certain density matrix ρ, the analogous expression is:

(320) $\left\langle I(\vec{x},t)\right\rangle = J\ \mathrm{Trace}\left(\hat{\rho}\,\hat{\vec{E}}^{-}(\vec{x},t)\hat{\vec{E}}^{+}(\vec{x},t)\right)$.

§25. Quantum theory of Young's interference experiment and the quantum degree of first-order coherence

In this section, we sketch the quantum theory of Young's interference experiment (from the point of view of the previously-developed formalism for the quantized electromagnetic field) and derive an expression for the so-called *quantum degree of first-order coherence*. See the figure below. Note



Young's two-slit experiment. A source of chaotic light is placed one focal length away from a lens, such that an approximately plane wave strikes the surface of a double slit. The intensity of the resulting interference pattern is recorded on a screen, shown at the right of the diagram. Position vectors of the two pinholes are $\vec{x}_1$ and $\vec{x}_2$ respectively. Lastly, s_1 and s_2 respectively denote the distance from the first pinhole to a given observation point on the screen, and the distance from the second pinhole to the same observation point on the screen.

theory, in the context of a standard calculation which culminates in the so-called "Fermi golden rule". See any elementary text on quantum mechanics, if this point is obscure to you and you would like a little more justification for why we are concerned with the matrix element given in equation (316).

[75] This argument can be made rigorous with time-dependent perturbation theory. A more sophisticated treatment of the quantum theory of the photo-electric detection of light based on such an analysis is given in Chapter 14 of L.Mandel and E.Wolf, *Optical Coherence and Quantum Optics*. Note, also, that for those studying Advanced Quantum Mechanics, time-dependent perturbation theory will be briefly reviewed near the beginning of that course.

[76] See footnote in the sentence immediately preceding equation (316).

[77] Note that this is a commonly used construct in quantum mechanics.

that, in the following, we assume a single polarization for simplicity and therefore drop the vector symbol from the electric field operators.

The positive-frequency part of the electric-field operator at a position $\vec{x}$ on the screen of a Young's two-slit experiment is:

$$\hat{E}^{+}(\vec{x},t) = u_1\hat{E}^{+}(\vec{x}_1,t_1) + u_2\hat{E}^{+}(\vec{x}_2,t_2), \tag{321}$$

where u_1 and u_2 are geometrix factors depending on the manner in which the radiation is diffracted by each of the slits. We are interested in the expectation value of the intensity operator for this experiment. Since we shall be dealing with mixed-state quantum fields described by a given density matrix, we need to calculate the expectation value of the intensity operator using result (320). Let us therefore adopt the notation:

$$\left\langle \hat{E}^{-}(\vec{x},t)\hat{E}^{+}(\vec{x},t)\right\rangle \equiv \mathrm{Trace}\left(\hat{\rho}\,\hat{E}^{-}(\vec{x},t)\hat{E}^{+}(\vec{x},t)\right). \tag{322}$$

Using (321) and expression (319) for the intensity operator, we calculate the expectation value of the intensity for Young's double-slit experiment:

$$\begin{aligned}\bar{I}(\vec{r}) &= J\left\langle \hat{E}^{-}(\vec{x},t)\hat{E}^{+}(\vec{x},t)\right\rangle \\ &\text{(now use (321))} \\ &= J\left\langle \begin{matrix}\left(u_1^{*}\hat{E}^{-}(\vec{x}_1,t_1)+u_2^{*}\hat{E}^{-}(\vec{x}_2,t_2)\right)\\ \times\left(u_1\hat{E}^{+}(\vec{x}_1,t_1)+u_2\hat{E}^{+}(\vec{x}_2,t_2)\right)\end{matrix}\right\rangle\end{aligned} \tag{323}$$

$$\begin{aligned}\bar{I}(\vec{r}) &= J\left\langle u_1^{*}\hat{E}^{-}(\vec{x}_1,t_1)u_1\hat{E}^{+}(\vec{x}_1,t_1)\right\rangle \\ &+ J\left\langle u_2^{*}\hat{E}^{-}(\vec{x}_2,t_2)u_2\hat{E}^{+}(\vec{x}_2,t_2)\right\rangle \\ &+ J\left\langle u_1^{*}\hat{E}^{-}(\vec{x}_1,t_1)u_2\hat{E}^{+}(\vec{x}_2,t_2)\right\rangle \\ &+ J\left\langle u_2^{*}\hat{E}^{-}(\vec{x}_2,t_2)u_1\hat{E}^{+}(\vec{x}_1,t_1)\right\rangle \\ &= \bar{I}_1(\vec{x}) + \bar{I}_2(\vec{x}) \\ &\quad + 2J\,\mathrm{Re}\left\langle u_1^{*}\hat{E}^{-}(\vec{x}_1,t_1)u_2\hat{E}^{+}(\vec{x}_2,t_2)\right\rangle\end{aligned} \tag{324}$$

Each of the first two terms (on the right side of the last line of this equation) denote the intensity of one pinhole when the other is blocked off. Therefore, *interference effects are described by the third term*. We see that the presence or absence of interference effects in Young's double slit experiment, is intimately related to the *correlation function* which evidently comprises this third term.

We introduce the *normalized quantum degree of first-order coherence:*

$$g^{(1)}(\vec{x}_1,t_1,\vec{x}_2,t_2) = \frac{\left\langle \hat{E}^{-}(\vec{x}_1,t_1)\hat{E}^{+}(\vec{x}_2,t_2)\right\rangle}{\sqrt{\left\langle \hat{E}^{-}(\vec{x}_1,t_1)\hat{E}^{+}(\vec{x}_1,t_1)\right\rangle}\sqrt{\left\langle \hat{E}^{-}(\vec{x}_2,t_2)\hat{E}^{+}(\vec{x}_2,t_2)\right\rangle}}, \tag{326}$$

where angular brackets are as denoted in (322). We rewrite result (323) in terms of our new notation:

$$\begin{aligned}\bar{I}(\vec{x}) &= \bar{I}_1(\vec{x}) + \bar{I}_2(\vec{x}) \\ &+ \left(\begin{matrix} 2J\,u_1^{*}u_2 \\ \times\sqrt{\left\langle E^{-}(\vec{x}_1,t_1)E^{+}(\vec{x}_1,t_1)\right\rangle} \\ \times\sqrt{\left\langle E^{-}(\vec{x}_2,t_2)E^{+}(\vec{x}_2,t_2)\right\rangle} \\ \times \mathrm{Re}\,g^{(1)}(\vec{x}_1,t_1,\vec{x}_2,t_2)\end{matrix}\right) \\ &= \bar{I}_1(\vec{x}) + \bar{I}_2(\vec{x}) \\ &\quad + 2\sqrt{\bar{I}_1(\vec{x})}\sqrt{\bar{I}_2(\vec{x})}\,\mathrm{Re}\,g^{(1)}(\vec{x}_1,t_1,\vec{x}_2,t_2).\end{aligned} \tag{327}$$

§26. Quantum degree of second-order coherence

We may define the normalised quantum degree of second-order coherence by the expression, which is a higher-order correlation function than that which was introduced in equation (326):

$$g^{(2)}(\vec{x}_1,t_1,\vec{x}_2,t_2;\vec{x}_2,t_2,\vec{x}_1,t_1) = \frac{\left\langle \hat{E}^{-}(\vec{x}_1,t_1)\hat{E}^{-}(\vec{x}_2,t_2)\hat{E}^{+}(\vec{x}_2,t_2)\hat{E}^{+}(\vec{x}_1,t_1)\right\rangle}{\left\langle \hat{E}^{-}(\vec{x}_1,t_1)\hat{E}^{+}(\vec{x}_1,t_1)\right\rangle\left\langle \hat{E}^{-}(\vec{x}_2,t_2)\hat{E}^{+}(\vec{x}_2,t_2)\right\rangle}. \tag{329}$$

Again, angular brackets are evaluated using the density matrix formalism as denoted in (322).

§27. Quantum coherence functions of arbitrary order and a criterion for a quantized electro-magnetic field to be coherent to all orders

Equations (326) and (329) are special cases of the normalized quantum degree of r^{th}-order coherence, which is a multi-point quantum correlation function of the following form:

$$
\begin{aligned}
&g^{(r)}\left(\vec{x}_1,t_1,...,\vec{x}_r,t_r;\vec{x}_{r+1},t_{r+1},...,\vec{x}_{2r},t_{2r}\right)\\
&=\frac{\left\langle\prod_{i=1}^{r}\hat{E}^{-}\left(\vec{x}_i,t_i\right)\prod_{j=r+1}^{2r}\hat{E}^{+}\left(\vec{x}_j,t_j\right)\right\rangle}{\prod_{i=1}^{2r}\sqrt{\left\langle\hat{E}^{-}\left(\vec{x}_i,t_i\right)\hat{E}^{+}\left(\vec{x}_i,t_i\right)\right\rangle}}\\
&=\frac{\mathrm{Trace}\left(\hat{\rho}\prod_{i=1}^{r}\hat{E}_i^{-}\left(\vec{x}_i,t_i\right)\prod_{j=r+1}^{2r}\hat{E}^{+}\left(\vec{x}_j,t_j\right)\right)}{\prod_{i=1}^{2r}\sqrt{\mathrm{Trace}\left(\hat{\rho}\,\hat{E}^{-}\left(\vec{x}_i,t_i\right)\hat{E}^{+}\left(\vec{x}_i,t_i\right)\right)}}
\end{aligned}
\tag{330}
$$

Once the density matrix associated with a given statistical quantum system is known, the infinite hierarchy of quantum correlation functions in equation (330) gives a means of calculating its associated degree of coherence to any order.

We consider a field to be (first-order) coherent if $|g^{(1)}|=1$, corresponding to *maximal fringe visibility* in the context of the Young's double-slit experiment. A field is second-order coherent if $|g^{(1)}|= g^{(2)}=1$.

Now, we now have a definition for the quantum degree of coherence of arbitrary order. We may therefore speak of a field as being *coherent to order "r"* if:

$$\left|g^{(r)}\right|=1,\quad r=1,2,\cdots,r\,. \tag{331}$$

One may define the quantized electromagnetic field to be *coherent to all orders* if the normalized quantum degree of r^{th}-order coherence has modulus unity for *all* values of r:

$$\left|g^{(r)}\right|=1,\quad r=1,2,\cdots,\infty\,. \tag{332}$$

With reference to (330), this implies that the un-normalized correlation function of order "r" has the following factorization property:

$$
\begin{aligned}
&\left|\left\langle\prod_{i=1}^{r}\hat{E}_i^{-}\left(\vec{x}_i,t_i\right)\prod_{j=r+1}^{2r}\hat{E}^{+}\left(\vec{x}_j,t_j\right)\right\rangle\right|\\
&=\prod_{i=1}^{2r}\sqrt{\left\langle\hat{E}^{-}\left(\vec{x}_i,t_i\right)\hat{E}^{+}\left(\vec{x}_i,t_i\right)\right\rangle}
\end{aligned}
\tag{333}
$$

Conversely, it is evident that *any* un-normalized correlation function:

$$\left\langle\prod_{i=1}^{r}\hat{E}_i^{-}\left(\vec{x}_i,t_i\right)\prod_{j=r+1}^{2r}\hat{E}^{+}\left(\vec{x}_j,t_j\right)\right\rangle \tag{334}$$

which is able to be factorized as a product of functions of the form:

$$
\begin{aligned}
&\left\langle\prod_{i=1}^{r}\hat{E}_i^{-}\left(\vec{x}_i,t_i\right)\prod_{j=r+1}^{2r}\hat{E}^{+}\left(\vec{x}_j,t_j\right)\right\rangle\\
&=\prod_{i=1}^{r}\Omega^{*}\left(\vec{x}_i,t_i\right)\prod_{j=r+1}^{2r}\Omega\left(\vec{x}_j,t_j\right)
\end{aligned}
\tag{335}
$$

will correspond to a quantized electromagnetic field which is coherent to all orders. (Note that the function Ω is independent of the index "i".)

Indeed, the *factorization property of the un-normalized correlation function* (335) is, in more formal treatments of this subject, adopted as the *definition* of coherence to all orders in a given pure-state or mixed-state quantized electromagnetic field. This factorizability of correlation functions, as a criterion for coherence, also appears in other areas such as theoretical condensed matter physics (see previous footnotes regarding this latter point).

§28. Coherent states are coherent to all orders

In this section we shall show that the *coherent states are coherent to all orders*, in the sense described in the previous article.

For a single-mode polarized field, equations (315a) and (315b) become[78]:

$$\hat{E}^{+}\left(\vec{x},t\right)\equiv i\sqrt{\frac{\hbar\omega}{2V}}\hat{a}e^{i\left(\vec{k}\bullet\vec{x}-\omega\,t\right)}\,, \tag{336a}$$

$$\hat{E}^{-}\left(\vec{x},t\right)\equiv -i\sqrt{\frac{\hbar\omega}{2V}}\hat{a}^{\dagger}e^{-i\left(\vec{k}\bullet\vec{x}-\omega t\right)}. \tag{336b}$$

Hence, for this case, the quantum degree of coherence (330) reduces to:

[78] We have dropped the vector symbol from the electric-field operators on the left-side of (336), and omitted the polarization vector from the right-side. This is consistent with our assumption of a *polarized* field.

$$g^{(r)}\left(\vec{x}_1,t_1,...,\vec{x}_r,t_r;\vec{x}_{r+1},t_{r+1},...,\vec{x}_{2r},t_{2r}\right)$$

$$=\frac{\left\langle\alpha\left|\prod_{i=1}^{r}\left(-i\sqrt{\frac{\hbar\omega}{2V}}\hat{a}^{\dagger}e^{-i\left(\vec{k}\bullet\vec{x}_i-\omega t_i\right)}\right)\times\prod_{j=r+1}^{2r}\left(i\sqrt{\frac{\hbar\omega}{2V}}\hat{a}e^{i\left(\vec{k}\bullet\vec{x}_j-\omega\, t_j\right)}\right)\right|\alpha\right\rangle}{\prod_{i=1}^{2r}\sqrt{\left\langle\alpha\left|-i\sqrt{\frac{\hbar\omega}{2V}}\hat{a}^{\dagger}e^{-i\left(\vec{k}\bullet\vec{x}_i-\omega t_i\right)}\times i\sqrt{\frac{\hbar\omega}{2V}}\hat{a}e^{i\left(\vec{k}\bullet\vec{x}_i-\omega\, t_i\right)}\right|\alpha\right\rangle}}$$

$$\left(\text{now cancel the common factors}\right)$$

$$=\frac{\left\langle\alpha\left|\left(\prod_{i=1}^{r}\hat{a}^{\dagger}e^{-i\left(\vec{k}\bullet\vec{x}_i-\omega t_i\right)}\right)\prod_{j=r+1}^{2r}\left(\hat{a}e^{i\left(\vec{k}\bullet\vec{x}_j-\omega\, t_j\right)}\right)\right|\alpha\right\rangle}{\prod_{i=1}^{2r}\sqrt{\left\langle\alpha\left|\hat{a}^{\dagger}e^{-i\left(\vec{k}\bullet\vec{x}_i-\omega t_i\right)}\hat{a}e^{i\left(\vec{k}\bullet\vec{x}_i-\omega\, t_i\right)}\right|\alpha\right\rangle}} \quad . \tag{337}$$

Now, return to the definitions (267a) of the coherent states, to allow us to proceed further:

$$g^{(r)}\left(\vec{x}_1,t_1,...,\vec{x}_r,t_r;\vec{x}_{r+1},t_{r+1},...,\vec{x}_{2r},t_{2r}\right)$$

$$=\frac{\left\langle\alpha\left|\prod_{i=1}^{r}\left(\alpha^{*}e^{-i\left(\vec{k}\bullet\vec{x}_i-\omega t_i\right)}\right)\prod_{j=r+1}^{2r}\left(\alpha\, e^{i\left(\vec{k}\bullet\vec{x}_j-\omega\, t_j\right)}\right)\right|\alpha\right\rangle}{\prod_{i=1}^{2r}\sqrt{\left\langle\alpha\left|\alpha^{*}e^{-i\left(\vec{k}\bullet\vec{x}_i-\omega t_i\right)}\alpha\, e^{i\left(\vec{k}\bullet\vec{x}_i-\omega\, t_i\right)}\right|\alpha\right\rangle}} \tag{338}$$

$$=\frac{\prod_{i=1}^{r}\left(\alpha^{*}e^{-i\left(\vec{k}\bullet\vec{x}_i-\omega t_i\right)}\right)\prod_{j=r+1}^{2r}\left(\alpha\, e^{i\left(\vec{k}\bullet\vec{x}_j-\omega\, t_j\right)}\right)\left\langle\alpha\,|\,\alpha\right\rangle}{\prod_{i=1}^{2r}\sqrt{\alpha^{*}e^{-i\left(\vec{k}\bullet\vec{x}_i-\omega t_i\right)}\alpha\, e^{i\left(\vec{k}\bullet\vec{x}_i-\omega\, t_i\right)}\left\langle\alpha\,|\,\alpha\right\rangle}}$$

(recall equation (276))

$$=\frac{\prod_{i=1}^{r}\left(|\alpha|e^{-i\theta}e^{-i\left(\vec{k}\bullet\vec{x}_i-\omega t_i\right)}\right)\prod_{j=r+1}^{2r}\left(|\alpha|e^{+i\theta}\, e^{i\left(\vec{k}\bullet\vec{x}_j-\omega\, t_j\right)}\right)}{\prod_{i=1}^{2r}|\alpha|}$$

$$=\frac{|\alpha|^{2r}\prod_{i=1}^{r}\left(e^{-i\left(\vec{k}\bullet\vec{x}_i-\omega t_i\right)}\right)\prod_{j=r+1}^{2r}\left(e^{i\left(\vec{k}\bullet\vec{x}_j-\omega\, t_j\right)}\right)}{|\alpha|^{2r}}$$

$$=\prod_{i=1}^{r}\left(e^{-i\left(\vec{k}\bullet\vec{x}_i-\omega t_i\right)}\right)\prod_{j=r+1}^{2r}\left(e^{i\left(\vec{k}\bullet\vec{x}_j-\omega\, t_j\right)}\right) \quad .$$

The modulus of this expression is unity for all values of r. Hence, according to (166), the coherent states are coherent to all orders. This is the ultimate justification for the use of the term "coherent state".

► *Exercise 51: We have just shown that the single-mode coherent states are coherent to all orders. This exercise examines the coherent properties of the multi-mode coherent states.*

Using the terminology developed in §18.3, with a slight but obvious change of notation, the multi-mode coherent states may be defined by:

$$\hat{a}_{\vec{k}_i}\left|\alpha_{\vec{k}_1}\alpha_{\vec{k}_2}\alpha_{\vec{k}_3}...\alpha_{\vec{k}_i}...\right\rangle=\alpha_{\vec{k}_i}\left|\alpha_{\vec{k}_1}\alpha_{\vec{k}_2}\alpha_{\vec{k}_3}...\alpha_{\vec{k}_i}...\right\rangle.$$

Using the definition (330) for the normalized quantum degree of r^{th}-order coherence, and expressions (315a) and (315b) for the positive- and negative-frequency components of the multi-mode electric field operator, please show that the multi-mode coherent states are coherent to all orders.

Please note: You should ignore the complications due to polarization (i.e. drop the vector symbol from the electric field operators on the left-hand side of (315a) and (315b), and omit the polarization vector from the right-hand side).

CHAPTER 5

ELEMENTS OF THE QUANTUM THEORY OF SCATTERING AND DIFFRACTION

§29. Free-space diffraction operator for Helmholtz fields[79]

Consider a time-independent complex scalar field $\psi(x,y,z)$, which obeys the so-called Helmholtz equation:

$$(\nabla^2+k^2)\psi(x,y,z)=0 \quad . \tag{338}$$

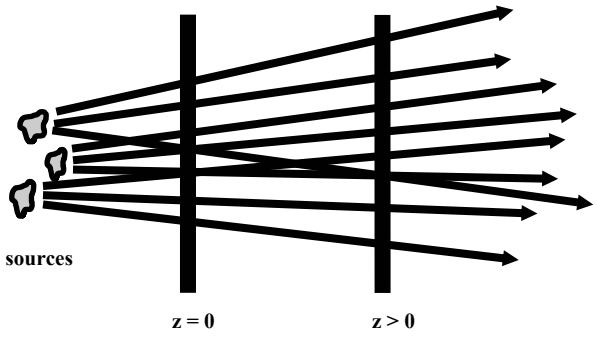
A field $\psi(x,y,z)$, which obeys (338), is said to be a "Helmholtz field".

[79] M.Nieto-Vesperinas, *Scattering and diffraction in physical optics (*Wiley, New York, 1991), section 2.2.

►*Exercise #52: Show that the following are Helmholtz fields: (a) Each component of the electric and magnetic field vector, for a free-space monoenergetic electromagnetic disturbance in a volume of space free of charges and currents, under an appropriate choice of gauge; (b) The spatial part of a free-space wavefunction which obeys the time-independent Schrödinger equation; (c) A stationary-state solution to the free-space Klein-Gordon equation, in the absence of either matter or electromagnetic fields; (d) Each component of a given Dirac wavefunction, which is a stationary-state solution to the free-space Dirac equation in the absence of charges and currents.*

In the present section, we study the free-space propagation of Helmholtz wavefields $\psi(x,y,z)$. As should be clear from the exercise above, this "diffraction problem" encompasses a wide class of both classical and quantum-mechanical wavefields.

With reference to the figure, suppose that all sources of the wavefield lie behind some plane located at z = 0. Given the complex value of the wavefield over the plane z = 0, we shall determine the propagated value of the wavefield over some plane z >0 downstream of the plane z = 0.



The diffraction problem

In our analysis of the free-space propagation of Helmholtz wavefields, we begin by considering the propagation of a very simple wavefield, namely the plane wave. Having solved this trivial propagation problem, we shall decompose a general Helmholtz wavefield as a sum of these elementary plane waves, each of which propagate trivially. Thus, we will have solved the diffraction problem for classical and quantum-mechanical Helmholtz wavefields by decomposing such objects into a weighted sum of simpler objects (elementary plane waves), and making use of the linearity of the Helmholtz equation in order to superpose such elementary plane waves in order to yield an arbitrary Helmholtz wavefield.

The elementary plane waves :

$$(339)\quad \psi_k^{(PW)}(x,y,z) = \exp\left(i\left(k_x x + k_y y + k_z z\right)\right)$$

are solutions of (338), provided that

$$(340)\quad k_x^2 + k_y^2 + k_z^2 = k^2 \Rightarrow k_z = \sqrt{k^2 - k_x^2 - k_y^2}\ .$$

We may not realize it yet, but in writing down (339) and (340), we have solved the trivial diffraction problem for the propagation of plane waves. To see this, let us evaluate our plane wave (339) over the plane z = 0 :

$$(341)\quad \psi_k^{(PW)}(x,y,z=0) = \exp\left(i\left(k_x x + k_y y\right)\right).$$

Given the value (341) of the plane-wavefield disturbance over the plane z = 0, we can obtain the *propagated* value of the plane-wave disturbance over any plane z > 0 which is downstream of z = 0 :

$$(342)\quad \begin{aligned}\psi_k^{(PW)}(x,y,z) &= \exp\left(i\left(k_x x + k_y y + k_z z\right)\right)\\ &= \exp\left(i\left(k_x x + k_y y\right)\right)\exp\left(iz\sqrt{k^2 - k_x^2 - k_y^2}\right).\end{aligned}$$

Having solved the trivial diffraction problem for elementary plane waves, we are ready to attack the diffraction problem for an wavefield which obeys either (338). We need to introduce the Fourier-integral representation of a two-dimensional function :

$$(343)\quad f(x,y) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} dk_x\, dk_y\ \hat{f}\left(k_x,k_y\right) \times \exp\left(+i\left(k_x x + k_y y\right)\right).$$

Thus we may express the wavefield disturbance in the plane z = 0 as a continuous sum of elementary plane waves (339) :

$$(344)\quad \psi(x,y,z=0) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} dk_x\, dk_y\ \hat{\psi}\left(k_x,k_y,z=0\right) \times \exp\left(+i\left(k_x x + k_y y\right)\right).$$

Now, we have already shown that each of the elementary plane-waves $\exp(i(k_x x + k_y y))$ in (341)

propagate trivially from z = 0 to z > 0 according to (342). Thus the propagated wavefield is given by the expression :

$$(345)\quad \psi(x,y,z)=\frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}dk_x\,dk_y\;\hat{\psi}(k_x,k_y,z=0)\times\exp\left(+i(k_x x+k_y y)\right)\times\exp\left(iz\sqrt{k^2-k_x^2-k_y^2}\right).$$

On comparison of (345) with (343), we write down the *angular-spectrum propagation law* :

$$(346)\quad \hat{\psi}(k_x,k_y,z)=\hat{\psi}(k_x,k_y,z=0)\times\exp\left(iz\sqrt{k^2-k_x^2-k_y^2}\right).$$

In words: the Fourier transform of the diffracted wavefield in the plane z > 0 is obtained by first taking the Fourier transform of the wavefield in the plane z = 0, and then multiplying by the propagation factor $\exp(iz\sqrt{(k^2-k_x^2-k_y^2)})$.

This powerful formalism is suitable for the calculation of diffraction patterns for both monochromatic electromagnetic waves and monoenergetic quantum-mechanical waves.

We close this section by writing (346) in operator form, whereby free-space propagation of a Helmholtz wavefield, from a plane $z = z$ to a plane $z = z + \Delta$ (where Δ is real), is obtained by applying the "diffraction operator" D_Δ to the unpropagated Helmholtz field, so that:

$$(347)\quad \begin{cases}\psi(x,y,z+\Delta)=D_\Delta\psi(x,y,z)\\ D_\Delta=F^{-1}\exp\left(i\Delta\sqrt{k^2-k_x^2-k_y^2}\right)F\end{cases}$$

where F denotes the Fourier transform with respect to x and y, under the convention of equation (343), and F^{-1} denotes the corresponding inverse Fourier transformation.

►*Exercise #53: Suppose we have a Helmholtz wave-field propagating from plane to plane, and suppose furthermore that its Fourier transform in the plane z = 0 is non-negligible only in a region* $k_x^2+k_y^2 << k^2$ *:*

$$(348)\quad \left|\hat{\psi}(k_x,k_y,z=0)\right|\approx 0 \text{ unless } k_x^2+k_y^2<<k^2.$$

Explain why this permits us to perform a binomial expansion on the square-root term in (345). Derive modified forms of both (345) and (346) under this binomial approximation. (NB : The resulting expressions are equivalent to a form of the famous "Fresnel diffraction integral".)

►*Exercise #54: Consider the propagation, through <u>time</u> rather than through space, of suitably-well-behaved but otherwise arbitrary solutions to the time-dependent free space Schrödinger equation. Utilizing the fact that such a quantum wave can be considered as a superposition of plane waves, show that solutions to this Schrödinger equation obey the following law for propagation through time :*

$$(349)\quad \Psi(\vec{r}_2,t_2)=\iiint K(\vec{r}_2-\vec{r}_1;t_2-t_1)\Psi(\vec{r}_1,t_1)d\vec{r}_1$$

where the propagator K has the following form :

$$(350)\quad K(\vec{\rho};\tau)=\frac{1}{h^3}\iiint\exp\left(\frac{i}{\hbar}(\vec{p}\bullet\vec{\rho}-E\tau)\right)d\vec{p}.$$

§30. Multislice formulation of electron scattering

Recall section §5.5, where we considered the non-relativistic limit of Dirac equation for an electron in an electromagnetic field. In equation (84) of that section, we obtained the Schrödinger-Pauli equation given below, for non-relativistic electrons in the presence of an electromagnetic field:

$$(351)\quad \begin{cases}H\Psi=E\Psi\\ H\equiv\frac{1}{2m}\left(-i\hbar\nabla-\frac{e\mathbf{A}}{c}\right)^2-\frac{e\hbar}{2mc}\boldsymbol{\sigma}\bullet\mathbf{B}+e\varphi\end{cases}.$$

We have restored the factors of $\hbar$ and c in the natural-units expression derived earlier, and work in Heaviside-Lorentz units. All other symbols are as defined in the Quantum Mechanics notes. Equation (20) is the non-relativistic Schrödinger-Pauli Hamiltonian for an electron in the presence of an electromagnetic field, with the said electromagnetic field being described by the electric potential φ and the magnetic field $\mathbf{B}=\nabla\times\mathbf{A}$.

Suppose, for simplicity, that there is no vector potential, and that the wavefunction is an energy eigenstate with energy E. Then:

$$(352)\quad \left(\nabla^2 + \frac{2mE}{\hbar^2}\left(1 - \frac{e\varphi}{E}\right)\right)\psi = 0 .$$

Now, it is easy to show that:

$$(353)\quad \frac{2mE}{\hbar^2} = \left(\frac{2\pi}{\lambda_0}\right)^2 ,$$

where λ_0 is the free-space de Broglie wavelength of the non-relativistic electron, in a region where the potential is zero. Hence (352) becomes:

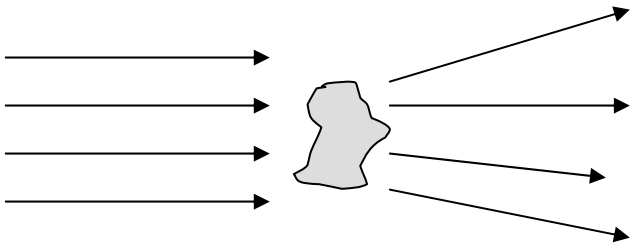
$$(354)\quad \left(\nabla^2 + \left(\frac{2\pi}{\lambda_0}\sqrt{1 - \frac{e\varphi}{E}}\right)^2\right)\psi = 0 .$$

In the absence of a vector potential, the term under the square root is unity, and λ_0 is equal to the free-space de Broglie wavelength. Now imagine that φ is everywhere equal to a non-zero constant, where $|\varphi| < E$. In this case, it is evident from (354) that the de-Broglie wavelength will now be:

$$(355)\quad \frac{2\pi}{\lambda} = \frac{2\pi}{\lambda_0}\sqrt{1 - \frac{e\varphi}{E}} \Rightarrow \lambda = \frac{\lambda_0}{\sqrt{1 - \frac{e\varphi}{E}}} .$$

If the electrostatic potential is a sufficiently slowly varying function of position, then the de Broglie wavelength in (355) will be a function of position.

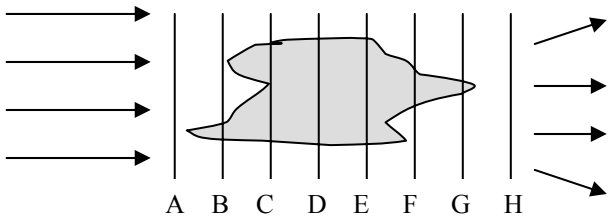
We now turn to our scattering problem, whereby a given monoenergetic electron is incident upon a region containing a given distribution of electrostatic potential φ, which scatters the incident beam. This scenario is sketched below.



A beam of non-relativistic electrons is incident from the left, upon a scatterer shown in grey. The electrostatic potential is nonzero, only within the grey volume; the magnetic vector potential is everywhere zero. The scattered electrons are shown on the right.

The electrostatic potential, which does the scattering, is considered to be sufficiently slowly varying for (355) to hold. Further, note that the requirement $|E| >> e\varphi$, which ensures that the incident electrons are sufficiently energetic to not be "captured" by the grey region in the figure above, ensures that the square root in (355) is always well defined.

Evidently, the evolution of the electron wavefunction, as it passes through the scattering region shown in grey, will be rather complicated. A "close up" of this process is shown below. Here,



we have a number of equally-spaced parallel planes A, B, ... , H, which are nominally perpendicular to the electron beam incident from the left. Let the spacing between each of these planes be Δ, and let the plane A coincide with z=0, where the z axis is parallel to the direction of the incident beam. We wish to calculate the so-called "exit surface wavefunction", which is the value of the electron wavefunction over the "exit surface" H, given the value of the electron wavefunction plane A.

By analogy with a loaf of bread which has been sliced, we can speak of the slab of space, between planes A and B, as a "slice". Thus, in the diagram above, the required scattering problem reduces to the problem of sequentially propagating the electron wavefunction through seven slices: from plane A to plane B, B to C, and so on, until we reach plane H.

Consider the slice of space, between the planes $z = m\Delta$, and $z = (m+1)\Delta$, where m is an integer. Suppose that the slice is sufficiently thin, for us to approximate φ, in the region anywhere within the slice, by the value φ takes over the plane $z = m\Delta$. Therefore, bearing (355) in mind, the *phase shift* $\delta\phi_{m\to m+1}(x, y)$ experienced by an approximately-z-directed electron, passing through the slice between the plane $z = m\Delta$ and the plane $z = (m+1)\Delta$, is

equal to 2π multiplied by the number of wavelengths that fit into the length Δ, so that:

$$(356)\quad \delta\phi_{m\to m+1}(x,y)=\frac{2\pi\Delta}{\lambda_0}\sqrt{1-\frac{e\varphi(x,y,z=m\Delta)}{E}}\,.$$

Unfortunately, (356) ignores the effects of the diffraction of the electron as it traverses the slice.

Bearing this criticism in mind, we are now ready to write down the multislice approximation for propagating the electron wavefield through a single slice:

$$(357)\quad \begin{aligned}&\psi(x,y,z=(m+1)\Delta)\\&=D_\Delta\left[\delta\phi_{m\to m+1}(x,y)\psi(x,y,z=m\Delta)\right]\end{aligned}$$

This formula is recursively applied to propagate the wave through the scattering region, from the entrance surface of the scattering volume, to the exit surface of the scattering volume[80]. Here, D_Δ is the free-space diffraction operator defined in equation (347), which is evidently applicable to (354) when $\varphi=0$. For obvious reasons, the algorithm just described in known as the "multislice algorithm", due to Cowley and Moodie.

The multislice algorithm assumes that propagating the electron wavefunction, through a given sufficiently thin slice, is the same as the following sequence of two steps: (i) change the phase of the wavefunction, as if the electron propagated through the slice in the absence of diffraction; (ii) then propagate through free space, through a distance equal to the thickness of the slice.

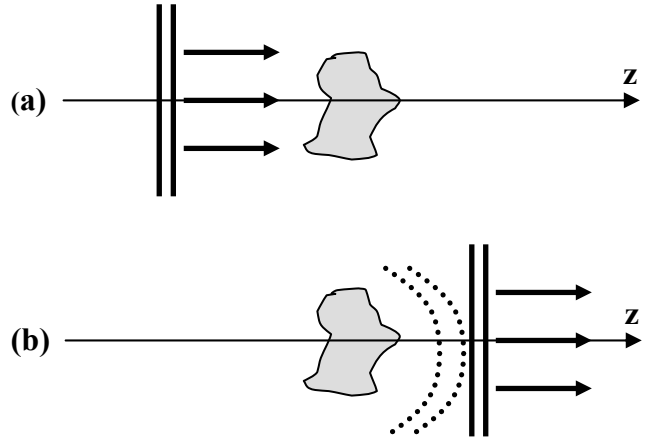
The multislice algorithm enjoys widespread use in the theory of electron diffraction. Note also, that the algorithm given here is easily modified in order to account for the interaction of the electron with a magnetic field, as must be done if the theory is to be applied to the electron microscopy of magnetic materials.

We close by pointing out that the ideas of multislice, as presented here, may be used outside electron optics, in contexts such as visible-light optics and x-ray optics.

[80] Note that, if one wants to propagate through free space beyond the exit surface, use can be made of (347) to propagate the exit-surface wavefunction through the required distance.

§31. Scattering, causality and analyticity[81]

The notion of causality, namely the demand that cause should always precede effect, has some profound and unexpected consequences in physics. Here we give a brief introduction to the role of causality in scattering theory. While our discussions will be restricted to the case of a classical complex time-dependent electromagnetic wave interacting with a scatterer, the calculations given here do have quantum mechanical counterparts which are beyond the scope of this course.



(a) A "δ-pulse" of classical light is incident upon the scattering material shown in grey. The pulse travels from right to left, at speed c; (b) After passing through the scattering region, the transmitted pulsed is a sum of the unscattered pulse (shown with a solid line) and the scattered pulse (shown with a dotted line).

Consider the scenario shown in the figure above, where a pulse of light is incident upon the scattering region shown in grey. The region outside the grey volume, is assumed to be free space. The pulse is only nonzero over the plane z=ct, which moves to the right at the speed of light. Assume that the grey region has a zero intersection with the half-space z<0, so that, before the time t=0, the pulse cannot yet have reached the scattering volume.

Therefore, respect for causality demands that the (time dependent) scattered disturbance, shown as the dotted lines in panel (b) of our figure, must

[81] J.J. Sakurai, *Advanced Quantum Mechanics*, Addison Wesley, Reading (167), pp. 59-64.

vanish when t<0. This statement of causality will now be shown to have important consequences on the functional form of the amplitude of the scattered radiation.

Via the Fourier representation of the Dirac delta:

$$\delta(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \exp(i\omega t)\,d\omega, \tag{358}$$

with t replaced by $(z/c) - t$, we conclude that the equation for the disturbance Ψ_0, due to the incident pulse, is:

$$\begin{aligned}\Psi_0(x,y,z,t) &= \delta\left(\frac{z}{c} - t\right) \\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty} \exp\left(i\omega\left(\frac{z}{c} - t\right)\right)d\omega\end{aligned} \tag{359}$$

Bearing in mind that $\omega/c = k$ for light, we see from the right side of (359) that the incident pulse is a linear superposition of z-directed monochromatic disturbances $\exp(ikz - i\omega t)$.

Since the equation for the field is assumed to be linear, the superposition principle holds. Accordingly, let us for the moment restrict ourselves to scattering by a single monochromatic disturbance $\exp(ikz - i\omega t)$. For large z, which corresponds to being sufficiently far from the (localized) scattering volume for it to appear pointlike, the scattered disturbance should look like a sum of the unscattered radiation, and a distorted spherical wave centered on the scattering volume. Thus we may write the following expression, valid for large z, for the wavefield Ψ_ω which results when a single monochromatic plane wave $\exp(ikz - i\omega t)$ is incident upon the scattering volume[82]:

$$\begin{aligned}\Psi_\omega \sim{}& \exp\left(i\omega\left(\frac{z}{c} - t\right)\right) \\ &+ f(\omega,\theta,\phi)\frac{1}{r}\exp\left(i\omega\left(\frac{r}{c} - t\right)\right),\end{aligned} \tag{360}$$

where f quantifies the deviations of the scattered wave from the "perfect" outgoing spherical wave:

$$\frac{1}{r}\exp\left(i\omega\left(\frac{r}{c} - t\right)\right), \tag{361}$$

and (r,θ,ϕ) denote spherical polar coordinates whose origin lies at a point on the z axis which is within the scattering volume.

Integrate equation (360) with respect to angular frequency, then divide by 2π and make use of (359), to give the following equation for the light-field $\Psi(x,y,z,t)$ which results when the delta-pulse in (359) interacts with the grey scattering potential:

$$\begin{aligned}\Psi(x,y,z,t) \sim{}& \delta\left(\frac{z}{c} - t\right) \\ &+ \frac{1}{2\pi r}\int_{-\infty}^{\infty} d\omega f(\omega,\theta,\phi)\exp\left(i\omega\left(\frac{r}{c} - t\right)\right).\end{aligned} \tag{362}$$

Let us now restrict ourselves to the forward scattering problem, which concerns itself with the value that $f(\omega,\theta,\phi)$ takes on the z axis. Denote this value by $f(\omega)$. On the z axis, sufficiently far to the right of the scattering volume, we therefore have:

$$\begin{aligned}\Psi(x=0,y=0,z,t) \sim{}& \delta\left(\frac{z}{c} - t\right) \\ &+ \frac{1}{2\pi z}\int_{-\infty}^{\infty} d\omega f(\omega)\exp\left(i\omega\left(\frac{z}{c} - t\right)\right).\end{aligned} \tag{363}$$

We are ready to invoke causality in the context of this scattering problem. The second term, on the right side of (363), which represents the scattered wave, must be zero during the period of time before the delta pulse has reach the scattering volume. However, we can go further than this: at a given time t, the scattered field must be zero in the half-space which lies *ahead* of the incident pulse, so that the scattered wave must be zero in the half space $z > ct$. Therefore:

[82] Note that a derivation of this expression, will be given in the next sub-section, in the context of quantum mechanical scattering. Since both monochromatic solutions to the free space d'Alembert equation, and monoenergetic solutions to the free-space time independent Schrödinger equation, are both Helmholtz fields, the quantum-mechanical derivation of the next subsequention may be used to derive the classical result in (360).

(364) $$\frac{1}{2\pi z}\int_{-\infty}^{\infty} d\omega f(\omega)\exp\left(i\omega\left(\frac{z}{c}-t\right)\right). = 0$$

when $z > ct$.

So, equation (364) vanishes when $z > ct$, i.e. when $(z/c) - t > 0$. Letting $-\tau \equiv (z/c) - t$, we see that the requirement $(z/c) - t > 0$ becomes $-\tau > 0$, which is the same as $\tau < 0$. In terms of τ, equation (364) becomes:

(365) $$\int_{-\infty}^{\infty} d\omega f(\omega)\exp(-i\omega\tau). = 0,\ \tau < 0.$$

This may be written as:

(366) $$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} d\omega f(\omega)\exp(-i\omega\tau) = \tilde{f}(\tau),$$
$$\tilde{f}(\tau) = 0 \text{ when } \tau < 0.$$

Therefore $\tilde{f}(\tau)$, namely the inverse Fourier transform of the forward-scattering amplitude $f(\omega)$, vanishes whenever $\tau < 0$.

Multiply both sides of (366) by $\exp(i\omega'\tau)$, then integrate both sides with respect to τ and make use of (358). Dropping the primes from the resulting expression, one arrives at:

(367) $$f(\omega) = \frac{1}{\sqrt{2\pi}}\int_{0}^{\infty} d\tau \tilde{f}(\tau)\exp(i\omega\tau).$$

Note that all negative frequencies are excluded from the integral on the right-hand side of (367), as a direct consequence of our demand for causality.

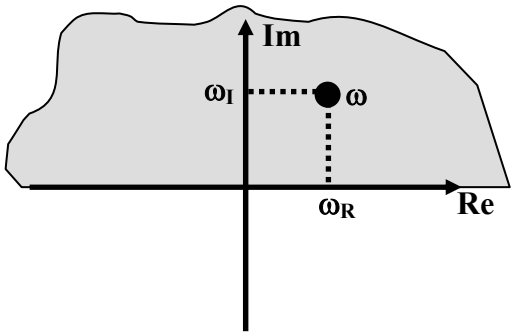
The title of this sub-section is "scattering, causality and analyticity" ... well, we have considered the first two terms in the title, so we now pass onto the third. The term "analyticity" is to be understood in the sense of analytic function, in complex analysis, which is essentially a calculus of complex functions of a single complex variable. I am aware that not all of you will have studied complex analysis, and so will make sure that all necessary complex-analysis results are stated explicitly in the text.

Equation (367) is a formula for $f(\omega)$, which will in general be a complex function f of a single real variable ω. We may "complexify" this expression, by replacing the real variable ω with the complex variable $\omega \equiv \omega_R + i\omega_I$, where ω_R and ω_I are both real numbers. Thus (367) becomes:

$$f(\omega_R + i\omega_I)$$

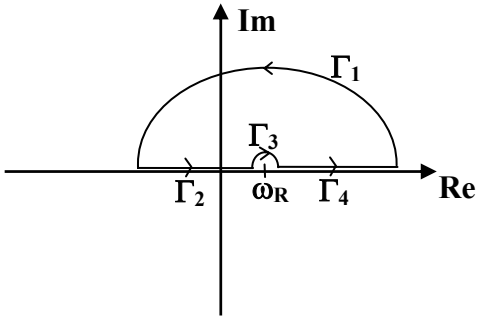
(368) $$= \frac{1}{\sqrt{2\pi}}\int_{0}^{\infty} d\tau \tilde{f}(\tau)\exp(i(\omega_R + i\omega_I)\tau) \quad .$$
$$= \frac{1}{\sqrt{2\pi}}\int_{0}^{\infty} d\tau \tilde{f}(\tau)\exp(i\omega_R\tau)\exp(-\omega_I\tau)$$

Evidently, $f(\omega_R + i\omega_I)$ is a complex function of a complex variable, which may be viewed as a "machine" taking a given complex number $\omega \equiv \omega_R + i\omega_I$ as input, and yielding another complex number $f(\omega_R + i\omega_I)$ as output.



Geometric view of $\omega \equiv \omega_R + i\omega_I$ as a number in the complex plane

We are now ready for some calculus in the complex plane. Consider the contour, in the complex plane, which is shown in the next figure. This closed contour consists of four connected parts: (i) Γ_1 is the upper half of a semi-circle, centered on the point ω_R on the real axis, which is traversed in an anti-clockwise sense and which has a radius R that will be made to tend to infinity from below; (ii) Γ_3 is the upper half of a semi-circle, centered on the point ω_R on the real axis, which is traversed in a



Path used for contour integral in the text

clockwise sense and which has a radius ε that will be made to tend to zero from above; (iii) Γ_2 is a line segment along the real axis, reaching from the point $\omega_R - R$ on the real axis, to the point $\omega_R - \varepsilon$ on the real axis, where R tends to infinity from below, and ε tends to zero from above; (iv) Γ_4 is a line segment along the real axis, which reaches from the point $\omega_R + \varepsilon$ on the real axis, to the point $\omega_R + R$ on the real axis, where R tends to infinity from below, and ε tends to zero from above.

The closed contour:

$$\Gamma = \Gamma_1 + \Gamma_2 + \Gamma_3 + \Gamma_4 \quad (369)$$

lies entirely in the so-called "upper half plane", which corresponds to the shaded grey region in the diagram on the previous page. Now, we saw that a mathematical consequence of our application of causality to the problem of the scattering of a delta-pulse of light from a localized potential, was that negative angular frequencies ω were absent from the integral in (367) ... when this expression was subsequently "complexified" in equation (368), the previously-mentioned absence of negative angular frequencies implies that the integrand in (368) is exponentially damped as $|\omega_R + i\omega_I| \to \infty$ in the upper half plane, due to the factor $\exp(-\omega_I \tau)$.

Now consider the following integral:

$$\oint_\Gamma \frac{f(\omega)d\omega}{\omega - \omega_R} = 0. \quad (370)$$

Our assertion, that the integral is zero, is a direct consequence of the fact that the integrand is a so-called "analytic function", everywhere on and inside the contour Γ. To those of you who have not studied complex analysis before, please take (370) as given, or see a complex analysis text!

Evidently, (370) can be written in the form:

$$\lim_{\substack{R\to\infty^- \\ \varepsilon\to 0^+}} \int_{\Gamma_1} \frac{f(\omega)d\omega}{\omega - \omega_R} + \lim_{\substack{R\to\infty^- \\ \varepsilon\to 0^+}} \int_{\Gamma_2+\Gamma_4} \frac{f(\omega)d\omega}{\omega - \omega_R} + \lim_{\varepsilon\to 0^+} \int_{\Gamma_3} \frac{f(\omega)d\omega}{\omega - \omega_R} = 0. \quad (371)$$

We now look at each integral separately.

First integral. The first integral, on the left side of (371), vanishes. To see this, note (from the last line of (368)) that $f(\omega_R + i\omega_I)$ vanishes exponentially rapidly as $|\omega_R + i\omega_I| \to \infty$ in the upper half plane.

Second integral. This is just an "ordinary" real integral, as follows:

$$\lim_{\substack{R\to\infty^- \\ \varepsilon\to 0^+}} \int_{\Gamma_2+\Gamma_4} \frac{f(\omega)d\omega}{\omega - \omega_R} = \lim_{\substack{R\to\infty^- \\ \varepsilon\to 0^+}} \left(\int_{\omega_R - R}^{\omega_R - \varepsilon} + \int_{\omega_R + \varepsilon}^{\omega_R + R} \right) \frac{f(\omega)d\omega}{\omega - \omega_R} \quad (372)$$

$$\equiv P \int_{-\infty}^{\infty} \frac{f(\omega)d\omega}{\omega - \omega_R}.$$

Here, P stands for "Cauchy principal value", which is just a shorthand for the limiting process indicated above, with singular points in the integrand being excluded from the region of integration.

Third integral[83]. We evaluate this by transforming to polar coordinates in the complex plane, with radial distance ε and angle θ both being measured with respect to the point ω_R in the complex plane. Accordingly, let $\omega - \omega_R = \varepsilon \exp(i\theta)$, so that $d\omega / d\theta = d\omega / d\theta = i\varepsilon \exp(i\theta)$

$$\begin{aligned}
&\lim_{\varepsilon\to 0^+} \int_{\Gamma_3} \frac{f(\omega)d\omega}{\omega - \omega_R} \\
&= \lim_{\varepsilon\to 0^+} \int_{\theta=\pi}^{\theta=0} \frac{f(\omega_R + \varepsilon\exp(i\theta))}{\varepsilon\exp(i\theta)} \frac{d\omega}{d\theta} d\theta \\
&= \lim_{\varepsilon\to 0^+} \int_{\theta=\pi}^{\theta=0} \frac{f(\omega_R + \varepsilon\exp(i\theta))}{\varepsilon\exp(i\theta)} i\varepsilon\exp(i\theta) d\theta \quad (373) \\
&= \lim_{\varepsilon\to 0^+} \int_{\theta=\pi}^{\theta=0} f(\omega_R + \varepsilon\exp(i\theta)) i d\theta \\
&= if(\omega_R) \int_{\theta=\pi}^{\theta=0} d\theta = -i\pi f(\omega_R)
\end{aligned}$$

Having evaluated each of the three integrals in equation (371), we can now write down:

$$\begin{aligned}
&P \int_{-\infty}^{\infty} \frac{f(\omega)d\omega}{\omega - \omega_R} - i\pi f(\omega_R) = 0, \\
&\Rightarrow f(\omega_R) = \frac{1}{i\pi} P \int_{-\infty}^{\infty} \frac{f(\omega)d\omega}{\omega - \omega_R}.
\end{aligned} \quad (374)$$

The real and imaginary parts, respectively, lead to the following so-called "Hilbert transform" pair:

[83] This integral can also be evaluated, a little more rapidly, using one of the formulae of complex analysis.

$$(375)\quad \begin{cases} \mathrm{Re}\{f(\omega)\} = \dfrac{1}{\pi} P \displaystyle\int_{-\infty}^{\infty} \dfrac{\mathrm{Im}\{f(\omega')\}d\omega'}{\omega' - \omega} \\ \mathrm{Im}\{f(\omega)\} = -\dfrac{1}{\pi} P \displaystyle\int_{-\infty}^{\infty} \dfrac{\mathrm{Re}\{f(\omega')\}d\omega'}{\omega' - \omega} \end{cases}$$

We conclude that the real and imaginary parts of the forward scattering amplitude, are related to one another via Hilbert transforms, as a direct consequence of our demand for causality in the classical scattering problem considered in this section. Similar "dispersion relations" occur in many areas of classical and quantum physics[84].

►*Exercise #55: Obtain equations (375).*

►*Exercise #56: By modifying the preceding argument, by applying it to the function* $f(\omega)/\omega$*, and under the assumption that f(0)=0, show that:*

$$(376)\quad \mathrm{Re}\{f(\omega)\} = \frac{\omega}{\pi} P \int_{-\infty}^{\infty} \frac{\mathrm{Im}\{f(\omega')\}d\omega'}{\omega'(\omega' - \omega)}.$$

§32. Green functions in scattering theory[85]

In the previous section, we took equation (360) as a "given". Here, we show how this equation for a scattered Helmholtz field, may be derived from the time-independent Schrödinger equation. Furthermore, we shall give an explicit expression for the scattering amplitude f, under the so-called "first Born approximation" of quantum scattering theory.

As a starting point, take the time-independent Schrödinger equation for a time-independent wavefunction $\psi(\mathbf{x})$ in the presence of a localized[86] time-independent potential $V(x,y,z)$:

$$(377)\quad \left(-\frac{\hbar^2}{2m}\nabla^2 + V(\mathbf{x})\right)\psi(\mathbf{x}) = E\psi(\mathbf{x}).$$

This can be written in the form of an inhomogeneous[87] Helmholtz equation, namely:

$$(378)\quad \left(\nabla^2 + k^2\right)\psi(\mathbf{x}) = U(\mathbf{x})\psi(\mathbf{x}),$$

where we have introduced the symbols $k^2 = 2mE/\hbar^2$ and $U(x,y,z) = 2mV(\mathbf{x})/\hbar^2$. Note that, in the absence of a scattering potential $U(\mathbf{x})$, equation (378) reduces to the free-space Helmholtz equation, which has been studied in previous sections.

The right hand side of (378) may be viewed as a "source term" for the scattered field. A Green function $G(\mathbf{x},\mathbf{x}')$, for a given inhomogeneous linear partial differential equation such as (378), is by definition equal to a solution to the said equation when the source term is proportional to a Dirac delta placed at the point $\mathbf{x}=\mathbf{x}'$ (physically, this corresponds to a point scatterer placed at the point $\mathbf{x}=\mathbf{x}'$, in the present context). Therefore, a Green function $G(\mathbf{x},\mathbf{x}')$, for equation (278), is by definition equal to a solution to the following:

$$(379)\quad \left(\nabla^2 + k^2\right)G(\mathbf{x},\mathbf{x}') = -4\pi\delta(\mathbf{x} - \mathbf{x}').$$

Note that the multiplicative factor of -4π, on the right side of this expression, is there by convention.

The central idea behind the Green function is this: *If one knows the Green function for a given linear differential equation, then by definition one knows a solution to the said equation, when a single point scatterer is present. Now consider an arbitrary scatterer being present. Since this arbitrary scatterer can be viewed as a continuous sum of single-point scatterers, and the equation is linear, one need only superpose the solutions for each of these point scatterers, in order to obtain a "solution"[88] in the presence of the arbitrary scatterer.*

Bearing the above in mind, we can immediately write down a "solution" to (378), namely:

$$(380)\quad -\frac{1}{4\pi}\iiint G(\mathbf{x},\mathbf{x}')U(\mathbf{x}')\psi(\mathbf{x}')dx'dy'dz'$$

[84] Dispersion relations exist for real and imaginary parts of the refractive index in classical optics, real and imaginary parts of the magnetic susceptibility in the theory of ferromagnetism, real and imaginary parts of certain coherence functions in the classical and quantum theories of coherence, and so forth

[85] E. Merzbacher, *Quantum Mechanics (second edition)*, John Wiley & Sons, New York (1970), pp. 223-231.

[86] By "localised", we mean that there exists a sphere, centered on the origin of coordinates, outside which V(x,y,z) vanishes.

[87] If the right side were zero, we would call this a "homogeneous Helmholtz equation".

[88] This is put in inverted commas, since we shall see that the resulting expression is not a solution at all, but rather an integral equation, which is fully equivalent to the original differential equation, with a given set of boundary conditions.

►*Exercise #57: Show that (380) is a "solution" to equation (378). Why is the word "solution" in inverted commas? What would be a more precise terminology, for expression (380)?*

Note that we spoke of (380) as *a* "solution", not *the* "solution", to (378). The reason for this is that, while there are infinitely many "solutions" to (378), we are after the physical "solution" which obeys a series of required boundary conditions.

These boundary conditions[89] are of two types. Firstly, we require the "solution", in the absence of a scatterer, to correspond to the unscattered wave. Thus, if the incident beam is a z-directed plane wave $\exp(ikz)$, which solves (378) when its right side is zero, then we should write (378) as:

$$\psi(\mathbf{x}) = \exp(ikz) - \frac{1}{4\pi}\iiint G(\mathbf{x},\mathbf{x}')U(\mathbf{x}')\psi(\mathbf{x}')dx'dy'dz' \quad (381)$$

The second type of boundary conditions regards the form of the wavefunction infinitely far away from the scattering potential. For example, we demand that the scattered radiation should be streaming *away* from the scattering potential, when we are far away from the said scattering potential, rather than towards it. While this may seem like a trivial demand, we shall see that it must be made in order for us to arrive at an unambiguous solution to our scattering problem.

Two problems remain: (a) First, we must determine the functional form of the Green function G, in equation (381). (b) Having done this, we will need to solve the integral equation (381), in which the wavefunction is the unknown, using the so-called "first Born approximation". We sequentially deal with each of these two tasks.

(a) Determination of Green function for (381)

Take equation (379), and write it as:

$$\left(\nabla^2 + k^2\right)G(\mathbf{x}-\mathbf{x}') = -4\pi\delta(\mathbf{x}-\mathbf{x}') . \quad (382)$$

Note that we have replaced $G(\mathbf{x},\mathbf{x}')$ by $G(\mathbf{x}-\mathbf{x}')$, which amounts to the requirement that the disturbance $G(\mathbf{x},\mathbf{x}')$ scattered from a point scatterer which is located at position $\mathbf{x}=\mathbf{x}'$, should depend only on the relative coordinate $\mathbf{x}-\mathbf{x}'$. Setting $\mathbf{x}'=0$ in (382), for the purpose of evaluating our Green function, we evidently need to solve the following equation for G:

$$\left(\nabla^2 + k^2\right)G(\mathbf{x}) = -4\pi\delta(\mathbf{x}) . \quad (383)$$

Let us write $G(\mathbf{x})$ as a Fourier integral, so that:

$$G(\mathbf{x}) = \frac{1}{\left(\sqrt{2\pi}\right)^3}\iiint \tilde{G}(\mathbf{k}')\exp(i\mathbf{k}'\bullet\mathbf{x})dk'_x dk'_y dk'_z . \quad (384)$$

Also, recall the following integral representation of the three-dimensional Dirac delta:

$$\delta(\mathbf{x}) = \frac{1}{(2\pi)^3}\iiint \exp(i\mathbf{k}'\bullet\mathbf{x})dk'_x dk'_y dk'_z . \quad (385)$$

Substitute (384) and (385) into (383), to give:

$$\frac{\left(\nabla^2 + k^2\right)}{\left(\sqrt{2\pi}\right)^3}\iiint \tilde{G}(\mathbf{k}')\exp(i\mathbf{k}'\bullet\mathbf{x})dk'_x dk'_y dk'_z$$

$$= -4\pi\frac{1}{(2\pi)^3}\iiint \exp(i\mathbf{k}'\bullet\mathbf{x})dk'_x dk'_y dk'_z$$

$$\iiint \tilde{G}(\mathbf{k}')\left[\left(\nabla^2 + k^2\right)\exp(i\mathbf{k}'\bullet\mathbf{x})\right]dk'_x dk'_y dk'_z$$

$$= -\sqrt{\frac{2}{\pi}}\iiint \exp(i\mathbf{k}'\bullet\mathbf{x})dk'_x dk'_y dk'_z$$

$$\iiint \left[\tilde{G}(\mathbf{k}')\left(-|\mathbf{k}'|^2 + k^2\right)\right]\exp(i\mathbf{k}'\bullet\mathbf{x})dk'_x dk'_y dk'_z$$

$$= \iiint\left[-\sqrt{\frac{2}{\pi}}\right]\exp(i\mathbf{k}'\bullet\mathbf{x})dk'_x dk'_y dk'_z \quad (386)$$

In the last of these equations, the quantities in square brackets must be equal to one another, so:

$$\tilde{G}(\mathbf{k}') = \frac{\sqrt{2/\pi}}{|\mathbf{k}'|^2 - k^2} . \quad (387)$$

Hence (384) becomes:

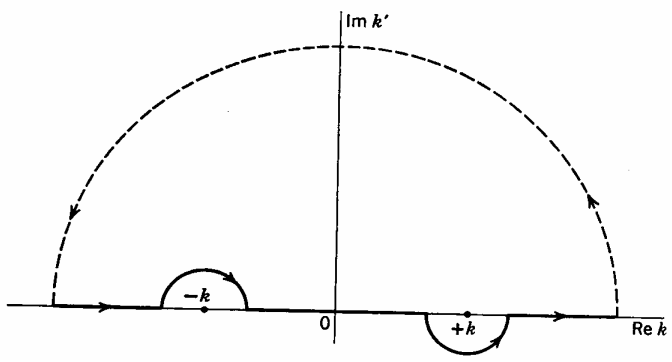
$$G(\mathbf{x}) = \frac{1}{2\pi^2}\iiint \frac{\exp(i\mathbf{k}'\bullet\mathbf{x})}{|\mathbf{k}'|^2 - k^2}dk'_x dk'_y dk'_z . \quad (388)$$

►*Exercise #58: Show that (388) can be re-written in the form:*

[89] We have been a little loose regarding the distinction between boundary values and boundary conditions. Please see me if you would like a more rigorous explanation than that given here.

$$(389)\quad G(\mathbf{x}) = -\frac{1}{\pi r}\frac{d}{dr}\int_{-\infty}^{\infty}\frac{\exp(ik'r)}{k'^2 - k^2}dk' .$$

As was the case in the previous section, for the purposes of evaluating this integral it will prove useful to consider k' as a complex variable, which may traverse the complex plane shown in the diagram below. Note that the integrand in (389), becomes undefined when $k' = \pm k$, which is why the contour, in the figure below, "hops around" these two points on the real axis.



Path of integration, in the complex k' plane, for calculation of outgoing Green function for the time-independent Schrödinger equation[90]

►*Exercise #59: By evaluating a suitable integral around the contour in the figure above, show that:*

$$(390)\quad G(\mathbf{x}) = \frac{\exp(ik\,|\,\mathbf{x}\,|)}{|\,\mathbf{x}\,|}$$

This Green function evidently obeys outgoing boundary conditions, as it represents a diverging wave when one is infinitely far away from the scatterer. This "outgoing" (or "retarded") Green function is the one appropriate to our problem. Alternative choices, for the contour above, lead to other Green functions.

Having determined our outgoing Green function, we write (381) as:

$$(391)\quad \psi(\mathbf{x}) = \exp(ikz) - \frac{1}{4\pi}\iiint\frac{\exp(ik\,|\,\mathbf{x}-\mathbf{x}'\,|)}{|\,\mathbf{x}-\mathbf{x}'\,|}U(\mathbf{x}')\psi(\mathbf{x}')dx'dy'dz'.$$

This is an integral equation, which we would like to solve for the wavefunction $\psi(x)$. To do so, we will use the first Born approximation.

(b) Solution to (391), with first Born approximation

The first Born approximation, which is valid when the scattering potential is sufficiently weak with respect to the incident plane wave $\exp(ikz)$, consists in replacing the scattered wavefield $\psi(\mathbf{x}')$, which appears in the integrand of (391), with the un-scattered field $\exp(ikz')$, so that:

$$(392)\quad \psi(\mathbf{x}) \approx \exp(ikz) - \frac{1}{4\pi}\iiint\frac{\exp(ik\,|\,\mathbf{x}-\mathbf{x}'\,|)}{|\,\mathbf{x}-\mathbf{x}'\,|}U(\mathbf{x}')\exp(ikz)dx'dy'dz'.$$

Regarding the denominator in this integral, let us assume that one is sufficiently far away from the scattering volume, so that one may replace $|\,\mathbf{x}-\mathbf{x}'\,|$ with $r \equiv \sqrt{x^2+y^2+z^2}$. Regarding the first exponent on the integrand of (392), we make the more sensitive first-order approximation:

$$(393)\quad \begin{aligned}|\mathbf{x}-\mathbf{x}'| &= \sqrt{|\,\mathbf{x}\,|^2 - 2\mathbf{x}\bullet\mathbf{x}' + |\,\mathbf{x}'\,|^2}\\ &\approx \sqrt{|\,\mathbf{x}\,|^2 - 2\mathbf{x}\bullet\mathbf{x}'} \equiv \sqrt{r^2 - 2\mathbf{x}\bullet\mathbf{x}'}\\ &= r\left(1 - \frac{2\mathbf{x}\bullet\mathbf{x}'}{r^2}\right)^{1/2} \approx r - \frac{\mathbf{x}\bullet\mathbf{x}'}{r}\end{aligned}$$

Hence (392) becomes:

$$(394)\quad \psi(\mathbf{x}) \approx \exp(ikz) - \frac{\exp(ikr)}{4\pi r}\iiint\exp\left(-\frac{ik\mathbf{x}\bullet\mathbf{x}'}{r}\right)U(\mathbf{x}')\exp(ikz')dx'dy'dz'.$$

Note that this has precisely the same form, which was written down in (360) on the grounds that it was intuitively reasonable. According to (394), the scattered wave $\psi(\mathbf{x})$ is equal to a sum of the unscattered wave exp(ikz), together with a modulated spherical wave which emerges from the origin of coordinates.

Let $\hat{\mathbf{x}} \equiv \mathbf{x}/r$ be a unit vector, so that (394) can be written in its final form:

$$(395)\quad \psi(\mathbf{x}) \approx \exp(ikz) + \frac{\exp(ikr)}{r}f(\hat{\mathbf{x}}),$$

where the scattering amplitude $f(\hat{\mathbf{x}})$ is given by:

[90] Taken from E. Merzbacher, *op cit.*, p. 226.

(396)

$$f(\hat{\mathbf{x}}) = \frac{-1}{4\pi}\iiint \exp(-ik\hat{\mathbf{x}} \bullet \mathbf{x}')U(\mathbf{x}')\exp(ikz')dx'dy'dz'$$

Note that the differential cross-section for the scattering process, denoted by $d\sigma / d\Omega$, is:

(397) $$\frac{d\sigma}{d\Omega} = |f(\hat{\mathbf{x}})|^2$$

►*Exercise #60: Consider scattering of a plane wave by the spherically-symmetric potential:*

(398) $$V(r) = \frac{V_0 \exp(-\alpha r)}{\alpha r},$$

where V_0 *and* α *are both constants, which are respectively equal to the strength of the potential, and the reciprocal of the range of the potential. Deduce the associated differential cross section, under the first Born approximation. Hence, or otherwise, deduce the cross section for scattering from a pure Coulomb potential. Note that your result should be in exact agreement with the famous Rutherford cross section.*