What is a Photon?

Foundations of Quantum Field Theory

C. G. Torre

June 16, 2018

What is a Photon?
Foundations of Quantum Field Theory

Version 1.0

PDF created June 16, 2018

Contents

Chapter 1

Introduction

This is a brief, informal, and relatively low-level course on the foundations of quantum field theory. The prerequisites are undergraduate courses in quantum mechanics and electromagnetism.

1.1 Why do we need this course?

I have always been dismayed by the fact that one can get a PhD in physics yet never be exposed to the theory of the photon. To be sure, we talk about photons all the time and we know some of their salient properties. But how do they arise in the framework of quantum theory? What do they have to do with the more familiar theory of electromagnetism? Of course, most people don't ever learn what a photon *really* is because a photon is an entity which arises in *quantum field theory*, which is not a physics class most people get to take. Quantum field theory typically arises in most physics curricula either in an advanced quantum mechanics course where one wants to do some many body theory, in a course on quantum optics, or in a course meant to explain the theory underlying high energy particle physics. Of course these subjects can be a bit daunting for someone who just wants to know what it is they are talking about when they use the term "photon". But the theory of the photon is not *that* complicated. The new ingredients needed to understand the theoretical description of the photon can be explained in some detail to anyone who has assimilated courses in quantum mechanics and electrodynamics. The principal goal of this course is really just to explain what are the main ideas behind the quantum field – this quantum "stuff" out of which everything is made. Constructing the quantum field in the context of electromagnetism leads immediately to the notion of the photon.

1.2 Why do we need quantum fields?

Let me remind you of the phenomenon of *spontaneous emission*, in which an atomic electron in an excited state will spontaneously emit one (or more) photons and end up in a lower energy state. This phenomenon will take place in the absence of any external stimulus and cannot be explained using the usual quantum mechanical model of the atom. Indeed, the usual quantum mechanical model of atomic energy levels represents them as stationary states. If an atomic electron occupying an atomic energy level were truly in a stationary state there could never be any spontaneous emission

since a stationary state has no time dependent behavior. The only way out of this conundrum is to suppose that atomic energy levels are not really stationary states once you take into account the interaction of photons and electrons. So now we need a quantum theory of more than one particle. We know how to do this, of course. But there is a wrinkle. Think again about the emission of a photon by an atomic electron. The initial state of the system has an electron. The final state of the system has an electron and a photon. Moreover, it is possible to have atomic transitions in which more than one photon appears/disappears. In the usual quantum mechanical formalism the number of particles is fixed. Indeed, the normalization of the wave function for a particle (or for particles) can be viewed as saying that the particle (or particles) is (or are) always somewhere. Clearly we will not be able to describe such processes using the standard quantum mechanical models. If this isn't enough, I remind you that there exist situations in which a photon may disappear, producing an electron-positron pair and, conversely, in which an electron-positron can turn into a photon. So even electrons and positrons are not immune from the appearance/disappearance phenomena.

Remarkably, it is possible to understand these variable-particle processes using the axioms of quantum theory, provided these axioms are used in a clever enough way. This new and improved use of quantum mechanics is usually called *quantum field theory* since it can be viewed as an application of the basic axioms of quantum mechanics to continuous systems (field theories) rather than mechanical systems. The picture that emerges is that the building blocks of matter and its interactions consist of neither particles nor waves, but a new kind of entity: a quantum field. Quantum mechanical particles and their wave-particle duality then become particularly simple manifestations of the quantum field. Every type of elementary particle is described by a quantum field (although the groupings by type depend upon the sophistication of the model). There is an electron field, a photon field, a neutrino field, a quark field and so forth.[1]

Quantum field theory (QFT) has led to spectacular successes in describing the behavior of a wide variety of atomic and subatomic phenomena. The success is not just qualitative; some of the most precise measurements known involve minute properties of the spectra of atoms. These properties are predicted via quantum field theory and, so far, the agreement with experiment is perfect.

Here I will give a brief, superficial introduction to some of the fundamental ideas underlying QFT. My principal goal will be just to show how QFT is used to describe photons and spontaneous emission. My strategy is to echo Dirac's original constructions, presented in a very influential monograph in 1930 [1], which was aimed at doing just this, and eventually led to modern quantum field theory. In doing so I have borrowed from Merzbacher's [2] and Sakurai's [3] treatment of some of these ideas, which are very clear and closely aligned with the concepts I want to introduce to you. It takes a bit more work to use these same ideas to describe, say, electrons in terms of quantum fields. Yet more work is needed to analyze the interaction of these particles (better: these fields). Quantum field theory is a very rich subject, still a subject of intense research.

1.3 Problems

1. What is the experimental value for the mean lifetime of the $2p$ state of hydrogen?

2. What is *positronium*? What is its mean lifetime?

[1] Relativistic fields include the anti-particles as well.

Chapter 2

The Harmonic Oscillator

The harmonic oscillator describes motion near stable equilibrium. Its quantum mechanical manifestation is precisely what is needed to understand free (non-interacting) fields and their particle excitations. So we must spend some time establishing the facts we will need concerning the quantum oscillator. This will also give me a chance to review with you the principal parts of quantum mechanics that we will need to define quantum field theory.

2.1 Classical mechanics: Lagrangian, Hamiltonian, and equations of motion

Before reviewing the quantum oscillator, it is good to first get oriented by reviewing its classical limit. A harmonic oscillator is characterized by 2 parameters: its mass m and its angular frequency ω. The Lagrangian and Hamiltonian for a (classical) harmonic oscillator are given respectively by

$$L(x,\dot{x}) = \frac{1}{2}m\dot{x}^2 - \frac{1}{2}m\omega^2 x^2, \quad H(x,p) = \frac{p^2}{2m} + \frac{1}{2}m\omega^2 x^2. \tag{2.1}$$

Here x is the displacement from equilibrium, $\dot{x}$ is the velocity, and p is the momentum canonically conjugate to x. The Euler-Lagrange equations are

$$\frac{\partial L}{\partial x} - \frac{d}{dt}\frac{\partial L}{\partial \dot{x}} = 0 \quad \Longleftrightarrow \quad \ddot{x} + \omega^2 x = 0. \tag{2.2}$$

The Hamilton equations are

$$\dot{x} = \frac{\partial H}{\partial p}, \quad \dot{p} = -\frac{\partial H}{\partial x} \quad \Longleftrightarrow \quad \dot{x} = \frac{p}{m}, \quad \dot{p} = -m\omega^2 x. \tag{2.3}$$

You can verify that the Lagrangian and Hamiltonian forms of the equations of motion are equivalent.

The general solution to the equations of motion is given by

$$x(t) = A\cos(\omega t) + B\sin(\omega t), \tag{2.4}$$

where

$$A = x(0), \quad B = \frac{1}{\omega}\dot{x}(0). \tag{2.5}$$

A very useful representation of the solutions to the oscillator equation involves complex amplitudes. Define

$$a = \sqrt{\frac{m\omega}{2}}(x + \frac{i}{m\omega}p), \quad a^* = \sqrt{\frac{m\omega}{2}}(x - \frac{i}{m\omega}p). \tag{2.6}$$

The variables a and a^* are just as good for describing the motion as x and p. In particular the displacement comes from the real part of a and the momentum from the imaginary part of a. In terms of the complex amplitudes, the general solution to the equation of motion takes the very simple form

$$a(t) = a(0)e^{-i\omega t}, \quad a^*(t) = a^*(0)e^{i\omega t}. \tag{2.7}$$

The Hamiltonian is also the energy of the oscillator. It is conserved, that is, unchanging in time:

$$H(t) = \frac{p(t)^2}{2m} + \frac{1}{2}m\omega^2 x(t)^2 = \frac{p(0)^2}{2m} + \frac{1}{2}m\omega^2 x(0)^2 = H(0). \tag{2.8}$$

In terms of the complex amplitudes this result is particularly simple:

$$H(t) = \omega a^*(t)a(t) = \omega \left(a^*(0)e^{i\omega t}\right)\left(a(0)e^{-i\omega t}\right) = \omega a^*(0)a(0) = H(0). \tag{2.9}$$

2.2 Classical mechanics: coupled oscillations

When a dynamical system has more than one degree of freedom, motion near stable equilibrium is in the form of *coupled harmonic oscillations*. Let us see how this happens.

Suppose the degrees of freedom are denoted by q^i, $i = 1, 2, \ldots, n$, and that the Lagrangian is of the form

$$L = \frac{1}{2}g_{ij}(q)\dot{q}^i\dot{q}^j - V(q). \tag{2.10}$$

I am using Einstein's *summation convention*; there is a double sum in the first term of L. The "metric" $g_{ij}(q)$ is an array which may depend upon the configuration coordinates. There is no loss of generality in assuming the metric to be symmetric,

$$g_{ij} = g_{ji}, \tag{2.11}$$

since only the symmetric combination appears in the sum over i and j. The metric involves the masses and other parameters; it is often just a collection of constants, but need not be. The metric is also often diagonal (*e.g.,* in spherical polar coordinates), but it need not be. The function V represents the potential energy of interaction among the degrees of freedom of the system and with their environment.

Critical points of V, *i.e.,* points q_0^i such that

$$\left(\frac{\partial V}{\partial q^i}\right)(q_0) = 0, \tag{2.12}$$

define equilibrium configurations of the system. We will suppose that the equilibrium is stable, *i.e.,* q_0^i is a (local) minimum of V. Let us approximate the motion in the neighborhood of a point of equilibrium by defining

$$y^i = q^i - q_0^i, \tag{2.13}$$

and expanding the Lagrangian in a Taylor series about $y^i = 0$. To the first non-trivial order we get

$$L \approx \frac{1}{2} M_{ij} \dot{y}^i \dot{y}^j - \frac{1}{2} K_{ij} y^i y^j \equiv L_2 \tag{2.14}$$

where

$$M_{ij} = g_{ij}(q_0), \qquad K_{ij} = \left(\frac{\partial^2 V}{\partial q^i \partial q^j} \right)(q_0), \tag{2.15}$$

and I have dropped an irrelevant additive constant $V(q_0)$, *i.e.,* the zero point of potential energy has been chosen to be at the equilibrium of interest, $q^i = q^i_0$. The array of constants $\mathbf{M}$ is symmetric, $M_{ij} = M_{ji}$, and we assume that the potential energy is sufficiently smooth so that the matrix $\mathbf{K}$ of second partial derivatives is symmetric at the critical point q_0:

$$K_{ij} = K_{ji}. \tag{2.16}$$

In this approximation, the Euler-Lagrange equations are

$$M_{ij} \ddot{y}^j + K_{ij} y^j = 0, \tag{2.17}$$

which are a coupled system of n homogeneous, linear ODEs with constant coefficients. Defining $\vec{y}$ as a column vector with entries y^i, and viewing M_{ij} and K_{ij} as symmetric matrices $\mathbf{M}$ and $\mathbf{K}$, we can write the EL equations in the matrix form:

$$\mathbf{M} \ddot{\vec{y}} = -\mathbf{K} \vec{y}. \tag{2.18}$$

Let us note that if q^i_0 is a point of stable equilibrium then the symmetric matrix $\mathbf{K}$ is non-negative, that is, it can have only non-negative eigenvalues.[1] This is because a negative eigenvalue will correspond to displacements y^i which lower the potential energy in an arbitrarily small neighborhood of the equilibrium point, which contradicts our assumption of stable equilibrium. If the eigenvalues are positive then the point q^i_0 is a local minimum. All this means that q^i_0 is a point of stable equilibrium if the *quadratic form*

$$K(\vec{y}) := K_{ij} y^i y^j \tag{2.19}$$

is positive, that is, $K(\vec{y}) > 0$ for all $\vec{y} \neq 0$. Physically this means that any displacement y^i from equilibrium will increase the potential energy. All this discussion is just restating standard results from multivariate calculus.

Likewise, positivity of the kinetic energy implies that the symmetric matrix $\mathbf{M}$ should be positive definite. This means the quadratic form

$$M(\vec{v}) := M_{ij} v^i v^j \tag{2.20}$$

is positive definite, *i.e.,* $M(\vec{v}) > 0$ for all $\vec{v} \neq 0$.

I now would like to appeal to a nice result from linear algebra. For any $\mathbf{K}$, if the quadratic form defined by $\mathbf{M}$ is positive definite, then there exists a linear change of coordinates $y^i \to x^i$:

$$x^i = \Lambda^i_j y_j, \quad \dot{x}^i = \Lambda^i_j \dot{y}^j, \tag{2.21}$$

[1] Note that a symmetric, real matrix always admits complete set of eigenvectors with real eigenvalues.

with inverse

$$y^i = (\Lambda^{-1})^i_j x^j, \quad \dot{y}^i = (\Lambda^{-1})^i_j \dot{x}^j. \tag{2.22}$$

such that in the new coordinates x^i the approximate Lagrangian takes the form

$$L_2 = \frac{1}{2}(\dot{x}_1^2 + \dot{x}_2^2 + \cdots + \dot{x}_n^2) - \frac{1}{2}(\omega_1^2 x_1^2 + \omega_2^2 x_2^2 + \cdots + \omega_n^2 x_n^2), \tag{2.23}$$

where ω_i, $i = 1, 2, \ldots, n$, are constants. In general these constants can be real, pure imaginary, or zero, depending on whether the eigenvalues of the matrix K are positive, negative, or zero, respectively. This result is called *simultaneous diagonalization of quadratic forms*. The new coordinates x^i given in (2.21) are called the *normal modes* of vibration. If the quadratic form K is non-negative then $\omega_i \geq 0$, and the constants ω_i are called the *characteristic frequencies*. I remind you that under a point transformation such as in (2.21), (2.22) the Euler-Lagrange equations for the new Lagrangian are the original Euler-Lagrange equations expressed in the new coordinates. The new form of the Lagrangian shows that near stable equilibrium the normal modes x^i oscillate independently and harmonically at the associated characteristic frequency $\omega_i > 0$.

To summarize: motion near stable equilibrium of any Lagrangian dynamical system can be viewed as a combination of independent simple harmonic oscillations in one dimension. This means that once you understand a single harmonic oscillator in one dimension (see the previous section) you, in principle, understand any system near stable equilibrium.

In the next section we will begin studying the quantum theory of a single harmonic oscillator. The result above suggests that the quantum behavior of more complicated systems near stable equilibrium can be described by a collection of such quantum oscillators, where the displacements correspond to the normal modes of vibration. This point of view leads to a successful description of a variety of physical phenomena, *e.g.,* vibrational spectra of molecules and phonons in solids. It is also one way of describing photons and, more generally, particle excitations of quantum fields.

2.3 The postulates of quantum mechanics

In the remainder of this chapter we will review the quantum theory of a simple harmonic oscillator. To begin, I should remind you what it means to have a "quantum theory" of some physical system. Here is a summary of the axioms of quantum mechanics.

- The "state" of a given physical system – the values of its observable properties at a given time – is represented by a unit vector in a (complex) Hilbert space $\mathcal{H}$. I will call this vector the "state vector". Recall that a Hilbert space is a vector space with an inner product which is Cauchy complete in the norm defined by that inner product. The vectors are denoted $|\psi\rangle$, the corresponding dual vectors defined by the inner product are denoted $\langle\psi|$, and the Hermitian (or "sesquilinear") inner product of $|\psi\rangle$ and $|\phi\rangle$ is denoted by $\langle\phi|\psi\rangle$. This scalar product is linear in $|\psi\rangle$ and anti-linear in $|\phi\rangle$.

- Any observable property A of a physical system is represented by a self-adjoint operator $\hat{A}$ on $\mathcal{H}$. The possible values that A can take are the (generalized) eigenvalues of $\hat{A}$. The eigenvectors of $\hat{A}$ represent states in which the corresponding eigenvalue will be measured with statistical certainty. More generally, the probability (density) of getting the (generalized) eigenvalue a in the state $|\psi\rangle$ is given by $|\langle a|\psi\rangle|^2$, where $|a\rangle$ is the (generalized) eigenvector of $\hat{A}$ with (generalized) eigenvalue a. If the eigenvalue is *degenerate* so that there is more than one

(generalized) eigenvector with the eigenvalue a then the probability is obtained by summing (integrating) $|\langle a|\psi\rangle|^2$ over the eigenspace. I will explain all this parenthetical "generalized" stuff in due course. For those who known about such things, I am allowing for self-adjoint operators with a continuous part to their spectra. Self-adjoint operators are used because their (generalized) eigenvectors will form a basis of $\mathcal{H}$. This is needed to make the probability interpretation work. Notice that any two state vectors differing by a phase, $|\psi\rangle$ and $e^{i\alpha}|\psi\rangle$, $\alpha \in \mathbf{R}$, have the same physical meaning.

- Time evolution of a physical system is represented by a 1-parameter family of state vectors $|\psi(t)\rangle$ defined by a self-adjoint operator, the *Hamiltonian* $\hat{H}$, via the Schrödinger equation:

$$\hat{H}|\psi(t)\rangle = i\hbar\frac{d}{dt}|\psi(t)\rangle. \tag{2.24}$$

- Given a splitting between the quantum system and its (classical) environment, a measurement of the the observable A of the quantum system by a measuring device in the environment with outcome a leaves the quantum system in the (generalized) state represented by $|a\rangle$. If the eigenvalue is degenerate, then one gets a (generalized) vector in the degenerate subspace.

If you are a seasoned quantum mechanic, these axioms will be old friends. If you are still getting proficient with quantum mechanics, the following illustration of the axioms via the quantum oscillator will hopefully help to get you where you need to be for this course.

2.4 The quantum oscillator

The Hilbert space for a harmonic oscillator can be chosen to be the set of complex functions on the real line which are square-integrable[2]

$$|\psi\rangle \quad \longleftrightarrow \quad \psi(x), \quad \int_{-\infty}^{\infty} dx\, |\psi(x)|^2 < \infty. \tag{2.25}$$

Here I am using the symbol $\leftrightarrow$ to mean "corresponds to". As you may know there are many equivalent ways to represent a given state vector. For example, one could just as well define $\mathcal{H}$ as the set of complex functions of $\psi(x)$ whose Fourier transform $\tilde{\psi}(k)$ is square integrable and then identify $\tilde{\psi}(k) \leftrightarrow |\psi\rangle$. Following Dirac's approach to quantum mechanics, different representations of the vectors are viewed as the expression of the vectors in different bases for $\mathcal{H}$. It is worth mentioning that the definition of the Hilbert space of square-integrable functions requires us to identify any function whose absolute-value-squared integrates to zero with the zero vector.

A Hilbert space has a scalar product. The scalar product of $|\psi\rangle$ and $|\phi\rangle$ is defined by

$$\langle\phi|\psi\rangle = \int_{-\infty}^{\infty} dx\ \phi^*(x)\psi(x). \tag{2.26}$$

This integral will exist for any square-integrable functions ϕ and ψ. States of the oscillator are to be identified with unit vectors:

$$1 = \langle\psi|\psi\rangle = \int_{-\infty}^{\infty} dx\, |\psi(x)|^2. \tag{2.27}$$

[2]The integration being used should be understood as Lebesgue integration, but this fine point won't matter much to us.

The linear operators $\hat{x}$ and $\hat{p}$ corresponding to the position[3] and momentum observables are defined by

$$\hat{x}|\psi\rangle \longleftrightarrow x\,\psi(x), \quad \hat{p}|\psi\rangle \longleftrightarrow \frac{\hbar}{i}\frac{d\psi}{dx}. \tag{2.28}$$

Here $\hbar$ is Planck's constant divided by 2π. I certainly will never prove such things here, but it is an important fact that these linear operators are *self-adjoint*, as required by the postulates. A necessary condition for this self-adjoint property is that the operators $\hat{x}$ and $\hat{p}$ are *symmetric*[4] with respect to the inner product. An operator $\hat{A}$ is symmetric if it is equal to its adjoint (on a suitable domain in $\mathcal{H}$). Recall that the adjoint of an operator $\hat{A}$ is the operator $\hat{A}^\dagger$ which satisfies

$$\langle\phi|\hat{A}^\dagger|\psi\rangle = \langle\psi|\hat{A}|\phi\rangle^* \tag{2.29}$$

for all vectors $|\phi\rangle$ and $|\psi\rangle$. A self-adjoint operator must satisfy $\hat{A} = \hat{A}^\dagger$, that is,

$$\langle\phi|\hat{A}|\psi\rangle = \langle\psi|\hat{A}|\phi\rangle^*. \tag{2.30}$$

Let us check that $\hat{x}$ is symmetric:

$$\langle\phi|\hat{x}|\psi\rangle = \int_{-\infty}^{\infty} dx\,\phi^*(x)x\psi(x) = \left(\int_{-\infty}^{\infty} dx\,\psi^*(x)x\phi(x)\right)^* = \langle\psi|\hat{x}|\phi\rangle^*. \tag{2.31}$$

It will be useful to note the *commutator* of the position and momentum operators. We have (on a suitable domain in $\mathcal{H}$)

$$[\hat{x},\hat{p}] \equiv \hat{x}\hat{p} - \hat{p}\hat{x} = i\hbar\hat{1}, \tag{2.32}$$

where $\hat{1}$ is the identity operator.

2.5 Energy spectrum

It turns out that the spectra of both $\hat{x}$ and $\hat{p}$ are continuous, so they have generalized eigenvectors and eigenvalues. I will defer the discussion of this until a bit later to keep things from getting too complicated too quickly. Instead, let me first focus on another observable: the Hamiltonian, or energy, H. The operator representing this observable is

$$\hat{H} = \frac{1}{2m}\hat{p}^2 + \frac{1}{2}m\omega^2\hat{x}^2, \tag{2.33}$$

where I am freely using the fact that one can add linear operators and multiply linear operators by scalars to create new linear operators. It can be shown that $\hat{H}$ is self-adjoint and that $\hat{H}$ has a purely discrete spectrum. The eigenvalue problem for $\hat{H}$ is to solve

$$\hat{H}|\psi\rangle = E|\psi\rangle \quad \longleftrightarrow \quad -\frac{\hbar^2}{2m}u''(x) + \frac{1}{2}m\omega^2x^2u(x) = E\,u(x) \tag{2.34}$$

for a constant E and (normalizable) $u(x)$. The solution of the differential equation (2.34) can be found in any quantum mechanics text. It turns out that the eigenvalues are given by

$$E_n = (n+\frac{1}{2})\hbar\omega, \quad n = 0,1,2,\ldots. \tag{2.35}$$

[3]Here "position" means "displacement from equilibrium".

[4]In the physics literature "symmetric" is, unfortunately, usually called "Hermitian". To make matters worse, mathematicians often use "Hermitian" to refer to bounded, self-adjoint operators.

The "quantum number" n is the number of "quanta" of energy $\hbar\omega$ the oscillator has relative to its lowest energy state – or "ground state". The energy eigenvectors correspond to functions which are polynomials in x times a Gaussian in x. Explicitly:

$$|E_n\rangle \quad \longleftrightarrow \quad u_n(x) = \frac{1}{\sqrt{2^n n!}} \left(\frac{m\omega}{\pi\hbar}\right)^{\frac{1}{4}} e^{-m\omega x^2/2\hbar} h_n(\sqrt{\frac{m\omega}{\hbar}}x), \tag{2.36}$$

where $h_n(\xi)$ is the Hermite polynomial in ξ,

$$h_n(\xi) = (-1)^n e^{\xi^2} \frac{d^n}{d\xi^n} e^{-\xi^2}, \quad n = 0, 1, 2, \dots . \tag{2.37}$$

Examples of Hermite polynomials are

$$h_0(\xi) = 1, \quad h_1(\xi) = 2\xi, \quad h_2(\xi) = 4\xi^2 - 2. \tag{2.38}$$

The energy eigenvectors are orthonormal:

$$\langle E_n | E_m \rangle = \int_{-\infty}^{\infty} dx\, u_n^*(x) u_m(x) = \delta_{nm}. \tag{2.39}$$

Because $\hat{H}$ is self-adjoint its eigenvectors form an orthonormal basis for the Hilbert space $\mathcal{H}$. This means that any element of the Hilbert space can be written as a superposition of energy eigenvectors, that is, there exist complex constants c_n such that[5]

$$|\psi\rangle = \sum_{n=0}^{\infty} c_n |E_n\rangle. \tag{2.40}$$

Here $c_n = \langle E_n | \psi \rangle$ are the components of $|\psi\rangle$ in the basis of energy eigenvectors. Alternatively, we have the operator identity

$$\sum_k |E_k\rangle\langle E_k| = \hat{1}, \tag{2.41}$$

so that

$$|\psi\rangle = \left(\sum_k |E_k\rangle\langle E_k|\right) |\psi\rangle = \sum_k |E_k\rangle\langle E_k|\psi\rangle = \sum_k c_k |E_k\rangle. \tag{2.42}$$

The normalization of $|\psi\rangle$, coupled with the orthonormality of the basis $|E_n\rangle$, means

$$\langle \psi | \psi \rangle = 1 = \sum_{n=0}^{\infty} |c_n|^2. \tag{2.43}$$

In light of (2.40), knowing the sequence of complex numbers $\{c_0, c_1, c_2, \dots\}$ is the same as knowing the state vector $|\psi\rangle$. Given the energy eigenvectors, the Hilbert space of square-integrable function can thus be identified with the set of square-summable sequences of complex numbers.

According to the rules of quantum mechanics, we can interpret $|c_n|^2$ as the probability that the energy E_n is measured when the oscillator is in the state given by $|\psi\rangle$. The normalization condition

[5]The infinite series converges to $|\psi\rangle$ in the sense that the difference of $|\psi\rangle$ and the sequence of partial sums from the right hand side defines a vector with zero norm in the limit as $n \to \infty$.

(2.43) then guarantees that the probabilities for all possible outcomes of an energy measurement add up to one.

The complex amplitudes (5.99) for the oscillator can be viewed as a "classical limit" of operators which have an important meaning relative to the energy spectrum. These operators are $\hat{a}$ and $\hat{a}^\dagger$, defined by

$$\hat{a} = \sqrt{\frac{m\omega}{2\hbar}}(\hat{x} + \frac{i}{m\omega}\hat{p}), \quad \hat{a}^\dagger = \sqrt{\frac{m\omega}{2\hbar}}(\hat{x} - \frac{i}{m\omega}\hat{p}). \tag{2.44}$$

Note that I have inserted a convenient factor of $\frac{1}{\sqrt{\hbar}}$ into the definition of each operator relative to its classical counterpart. This redefinition makes the amplitudes dimensionless. Because $\hat{x}$ and $\hat{p}$ are self-adjoint, you can see from inspection that a and $a^\dagger$ are adjoints of each other. Using the commutation relations between position and momentum it is easy to see that

$$[\hat{a}, \hat{a}^\dagger] = \hat{1} \tag{2.45}$$

and

$$\hat{H} = \hbar\omega(\hat{a}^\dagger\hat{a} + \frac{1}{2}\hat{1}). \tag{2.46}$$

The operator $\frac{1}{2}\hbar\omega\hat{1}$ represents a simple shift of the zero point of energy from zero. One could simply drop this term; this amounts to measuring the energy of the oscillator relative to its ground state energy. The operator $\hat{N} \equiv \hat{a}^\dagger\hat{a}$ represents the *energy quanta* observable; it is also called the *number operator*. I say this since it turns out that

$$\hat{N}|E_n\rangle = n|E_n\rangle. \tag{2.47}$$

Notice in particular that the ground state satisfies

$$\hat{N}|E_0\rangle = 0, \quad \Longrightarrow \quad \langle E_0|\hat{a}^\dagger\hat{a}|E_0\rangle = 0. \tag{2.48}$$

The second equation says that the norm of the vector $\hat{a}|E_0\rangle$ vanishes, which also implies the first equation. Therefore the ground state is characterized by the condition

$$\hat{a}|E_0\rangle = 0. \tag{2.49}$$

More generally, it is easy to compute:

$$[\hat{N}, \hat{a}] = -\hat{a}, \quad [\hat{N}, \hat{a}^\dagger] = \hat{a}^\dagger. \tag{2.50}$$

This implies that $\hat{a}|E_n\rangle \propto |E_{n-1}\rangle$ and $\hat{a}^\dagger|E_n\rangle \propto |E_{n+1}\rangle$. Explicit computation reveals

$$\hat{a}|E_n\rangle = \sqrt{n}|E_{n-1}\rangle, \quad \hat{a}^\dagger|E_n\rangle = \sqrt{n+1}|E_{n+1}\rangle, \quad n = 1, 2, \ldots. \tag{2.51}$$

For this reason the operators $\hat{a}$, $\hat{a}^\dagger$ are sometimes called "ladder operators", or "annihilation and creation operators" of energy quanta. These operators, mathematically speaking, add and subtract energy quanta from energy eigenstates. It is possible to derive the entire energy spectrum – including the energy eigenfunctions – just using the commutation relations of $\hat{N}$, $\hat{a}$, and $\hat{a}^\dagger$. This means the algebra of the ladder operators more or less *defines* the quantum oscillator.

2.6 Position, momentum, and their continuous spectra

The mathematical model of operators representing observables with continuous spectra, *e.g.,* the position and momentum operators, is a little more complicated than that for operators representing observables with discrete spectrum such as the oscillator energy. Operators with discrete spectrum are in many ways like ordinary matrices, albeit with an infinite number of matrix elements. Operators with continuous spectrum require some new technology. Let us begin by seeing the difficulties which arise when trying to solve the eigenvalue problem for operators such as position and momentum.

2.6.1 Position

The eigenvalue problem for the operator representing position should be

$$\hat{x}|y\rangle = y|y\rangle \quad \longleftrightarrow \quad x\psi_y(x) = y\psi_y(x), \tag{2.52}$$

where y is some (hopefully real) number. No non-trivial function can satisfy this condition. Indeed, if the function is non-zero for at least two values of x there can be no solution. But if the function is non-zero for only one value of x then it is (equivalent to) the zero vector in the Hilbert space. It is possible to solve this position eigenvalue equation if one allows a more general type of eigenfunction, a "generalized function", the Dirac delta function $\delta(x, y)$. The solution of the eigenvalue equation is then written

$$\psi_y(x) = \delta(x, y). \tag{2.53}$$

We shall need some of this technology, so let us digress briefly to review it.

Strictly speaking, the Dirac delta function is not a function, but rather a *distribution.*[6] It is possible to treat the Dirac delta as a limit of suitable elements of the Hilbert space. The limit will not exist in the Hilbert space, but the limit of suitable scalar products will exist, and that is all we shall need since the probability interpretation only needs a way to define the scalar products $\langle y|\psi\rangle$ for all possible values of y. A standard example of this limiting process is as follows. Let y be some given real number and consider the following 1-parameter family of elements of $\mathcal{H}$:

$$|\phi_{\epsilon,y}\rangle \quad \longleftrightarrow \quad \phi_{\epsilon,y}(x) = \frac{1}{2\pi}\int_{-\frac{1}{\epsilon}}^{\frac{1}{\epsilon}} dk\, e^{ik(x-y)}, \quad \epsilon > 0. \tag{2.54}$$

It is straightforward to compute the integral and find

$$|\phi_{\epsilon,y}(x)|^2 = \frac{\sin^2\left(\frac{\pi}{\epsilon}(x-y)\right)}{\pi^2(x-y)^2}, \tag{2.55}$$

from which you can see that the function $\phi_{\epsilon,y}(x)$ is square integrable. As a nice exercise you can check that

$$\langle\phi_{\epsilon,y}|\phi_{\epsilon,y}\rangle = \int_{-\infty}^{\infty} dx\, |\phi_{\epsilon,y}(x)|^2 = \frac{1}{\epsilon}, \tag{2.56}$$

[6]Here, a distribution is a continuous linear function on a dense subspace $\mathcal{S} \subset \mathcal{H}$ of the Hilbert space, where "continuous" implies a choice of appropriate appropriate topology on $\mathcal{S}$. I shall generally avoid this more rigorous way of doing things.

showing rather clearly how this function fails to be square-integrable in the limit $\epsilon \to 0$. The *generalized eigenfunction* of position – the delta function – arises in the limit as $\epsilon \to 0$. Although we often state this result by writing

$$\delta(x,y) = \frac{1}{2\pi}\int_{-\infty}^{\infty} dk\, e^{ik(x-y)}, \tag{2.57}$$

this equation has to be treated carefully. What it means is that for vectors $|\chi\rangle$ in a suitable subspace $\mathcal{S} \subset \mathcal{H}$ the limit of $\langle\phi_{\epsilon,y}|\chi\rangle$ as $\epsilon \to 0$ is defined. We have

$$\langle\phi_{\epsilon,y}|\chi\rangle = \frac{1}{2\pi}\int_{-\infty}^{\infty} dx \int_{-\frac{1}{\epsilon}}^{\frac{1}{\epsilon}} dk\, e^{-ik(x-y)}\chi(x). \tag{2.58}$$

For suitably nice functions[7] one can interchange the orders of integration and then take the limit as $\epsilon \to 0$. After interchanging the order of integration we get

$$\langle\phi_{\epsilon,y}|\chi\rangle = \frac{1}{2\pi}\int_{-\frac{1}{\epsilon}}^{\frac{1}{\epsilon}} dk \int_{-\infty}^{\infty} dx\, e^{-ik(x-y)}\chi(x) = \frac{1}{2\pi}\int_{-\frac{1}{\epsilon}}^{\frac{1}{\epsilon}} dk\, \tilde{\chi}(k)e^{iky}, \tag{2.59}$$

where $\tilde{\chi}(k)$ is the *Fourier transform* of $\chi(x)$. As $\epsilon \to 0$ we get the Fourier representation of the function $\chi(y)$:

$$\lim_{\epsilon\to 0}\langle\phi_{\epsilon,y}|\chi\rangle = \frac{1}{2\pi}\int_{-\infty}^{\infty} dk\, \tilde{\chi}(k)e^{iky} = \chi(y). \tag{2.60}$$

So, we have

$$\lim_{\epsilon\to 0}\int_{-\infty}^{\infty} dx\, \phi^*_{\epsilon,y}(x)\chi(x) = \chi(y), \tag{2.61}$$

which is the defining relation for the delta function.

If desired, one can opt to avoid all this delta function stuff and always work with a small but non-vanishing ϵ. Then the position will not be defined as a mathematical point, but only in terms of a region of size determined by ϵ, but if ϵ is small enough it will not matter physically. While this is satisfying in that one does not need to introduce distributions, this framework is somewhat cumbersome, of course, and that is why one likes to set $\epsilon = 0$ and use the delta function.

The preceding results are usually packaged in various simple "equations". If you know what you are doing, these "equations" are very helpful. If you let the notation substitute for understanding, then eventually there will be trouble. The most common relations one uses are:

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} dk\, e^{ik(x-y)} = \delta(x,y) = \delta(y,x) \equiv \delta(x-y), \tag{2.62}$$

$$\int_{-\infty}^{\infty} dx\ \delta(x,y)f(x) = f(y), \tag{2.63}$$

$$f(x)\delta(x,y) = f(y)\delta(x,y), \tag{2.64}$$

[7]For our purposes, a class of "nice" functions would be smooth functions whose absolute value decreases faster than the reciprocal of any polynomial in x as $|x| \to \infty$.

$$\lim_{\epsilon\to 0} |\phi_{\epsilon,y}\rangle = |y\rangle, \tag{2.65}$$

$$|y\rangle \quad \longleftrightarrow \quad \delta(x,y), \tag{2.66}$$

$$\hat{x}|y\rangle = y|y\rangle, \tag{2.67}$$

$$\langle y|\chi\rangle = \chi(y), \tag{2.68}$$

$$\int_{-\infty}^{\infty} dy\, \chi(y)|y\rangle = |\chi\rangle, \tag{2.69}$$

$$\langle y|z\rangle = \delta(y,z), \tag{2.70}$$

$$\int_{-\infty}^{\infty} dx\, |x\rangle\langle x| = \hat{1}. \tag{2.71}$$

These last five formal relations are the most important for us. If we augment our Hilbert space by including generalized functions like the delta function then the eigenvectors of position can be accommodated and they constitute a generalization of an orthonormal basis, now labeled by a continuous variable instead of a discrete variable, with summations over eigenvectors becoming integrals, and with the Dirac delta function replacing the Kronecker delta in the orthonormality relation.

It will be useful later to have available the following relations between the energy eigenvectors and position eigenvectors:

$$u_k(x) = \langle x|E_k\rangle. \tag{2.72}$$

$$\int_{-\infty}^{\infty} dx\, u_k^*(x)u_l(x) = \int_{-\infty}^{\infty} dx\, \langle E_k|x\rangle\langle x|E_l\rangle = \langle E_k|E_l\rangle = \delta_{kl}, \tag{2.73}$$

$$\sum_k u_k^*(x)u_k(y) = \sum_k \langle y|E_k\rangle\langle E_k|x\rangle = \langle y|x\rangle = \delta(x,y). \tag{2.74}$$

Finally, let me remind you of the physical interpretation of the (position) wave function,

$$|\psi\rangle \longleftrightarrow \psi(x) = \langle x|\psi\rangle. \tag{2.75}$$

In the state defined by $\psi(x)$, the probability $P_x(a,b)$ for finding the position $x \in (a,b)$ is given by

$$P_x(a,b) = \int_a^b dx\, |\psi(x)|^2. \tag{2.76}$$

We say that $|\psi(x)|^2$ is the *probability density* for position, and that $|\psi(x)|^2 dx$ is the probability for finding the particle in the infinitesimal interval $(x, x+dx)$.

2.6.2 Momentum

For the momentum operator we want to solve an eigenvalue problem of the form

$$\hat{p}|p\rangle = p|p\rangle \quad \longleftrightarrow \quad \frac{\hbar}{i}\frac{d}{dx}u_p(x) = pu_p(x), \tag{2.77}$$

where p is a real constant. It is not too hard to solve this differential equation; we have

$$u_p(x) = Ae^{\frac{i}{\hbar}px}, \tag{2.78}$$

where A is a constant. There are two difficulties here. First of all, the spectrum of the momentum operator should consist of real numbers, but the eigenvalue equation allows p to be complex. Secondly, and perhaps more drastically, whether p is real or complex the functions $u_p(x)$ are not square-integrable – they are not elements of the Hilbert space. We are in a similar place as we were with the position operator, and we can proceed in a similar way. Once again we can define the momentum eigenfunctions as generalized functions (or "distributions") via limits of elements of $\mathcal{H}$, with the proviso that these limits are only used in scalar products with suitable elements of $\mathcal{H}$. The details are as follows.

Fix a real number p and consider a 1-parameter family of vectors in $\mathcal{H}$:

$$|\mu_{p,\epsilon}\rangle \quad \longleftrightarrow \quad u_{p,\epsilon}(x) = \frac{1}{\sqrt{2\pi\hbar}}e^{\frac{i}{\hbar}px}e^{-\epsilon^2x^2}, \quad \epsilon > 0. \tag{2.79}$$

It is easy to see that these functions are square-integrable for any $\epsilon > 0$:

$$\langle\mu_{p,\epsilon}|\mu_{p,\epsilon}\rangle = \frac{1}{4\hbar}\sqrt{\frac{2}{\pi}}\frac{1}{\epsilon}. \tag{2.80}$$

Of course, $u_{p,\epsilon}(x)$ is not a momentum eigenfunction, but it does approach the "generalized" eigenfunction (2.78) as $\epsilon \to 0$. While the limit does not lead to an element of $\mathcal{H}$, scalar products with elements of $\mathcal{H}$ do admit a limit. We have

$$\langle\mu_{p,\epsilon}|\psi\rangle = \frac{1}{\sqrt{2\pi\hbar}}\int_{-\infty}^{\infty} dx\, e^{-\frac{i}{\hbar}px}e^{-\epsilon^2x^2}\psi(x). \tag{2.81}$$

Evidently, as $\epsilon \to 0$ the limit of the scalar product is proportional to the Fourier transform $\tilde{\psi}$ of ψ:

$$\lim_{\epsilon\to 0}\langle\mu_{p,\epsilon}|\psi\rangle = \frac{1}{\sqrt{2\pi\hbar}}\int_{-\infty}^{\infty} dx\, e^{-\frac{i}{\hbar}px}\psi(x) = \frac{1}{\sqrt{\hbar}}\tilde{\psi}(k), \quad k = p/\hbar. \tag{2.82}$$

One usually uses a notation such as $\phi(p) = \frac{1}{\sqrt{\hbar}}\tilde{\psi}(k)$ and calls $\phi(p)$ the *momentum wave function*. Since knowing the Fourier transform of a function is as good as knowing the function, and since the Fourier transform $\tilde{\psi}$ is square-integrable if and only if ψ is, we could define $\mathcal{H}$ as the set of square-integrable momentum wave functions. This is the basis for using momentum space wave functions to "do" quantum mechanics.

As with the technology surrounding generalized position eigenvectors, the preceding results are packaged in various simple notations. The most common are:

$$\lim_{\epsilon\to 0}|\mu_{p,\epsilon}\rangle = |p\rangle \quad \longleftrightarrow \quad \lim_{\epsilon\to 0}u_{p,\epsilon}(x) \equiv u_p(x) = \frac{1}{\sqrt{2\pi\hbar}}e^{\frac{i}{\hbar}px}, \tag{2.83}$$

$$\hat{p}|p\rangle = p|p\rangle \tag{2.84}$$

$$\langle p|\psi\rangle = \frac{1}{\sqrt{\hbar}}\tilde{\psi}(p/\hbar) = \phi(p), \tag{2.85}$$

$$\langle x|p\rangle = \frac{1}{\sqrt{2\pi\hbar}}e^{\frac{i}{\hbar}px}, \quad \langle p|x\rangle = \frac{1}{\sqrt{2\pi\hbar}}e^{-\frac{i}{\hbar}px}, \tag{2.86}$$

$$\int_{-\infty}^{\infty} dp\, \phi(p)|p\rangle = |\psi\rangle. \tag{2.87}$$

$$\langle p'|p\rangle = \delta(p', p), \tag{2.88}$$

$$\int_{-\infty}^{\infty} dp\, |p\rangle\langle p| = \hat{1}. \tag{2.89}$$

There are relations involving the energy eigenvectors analogous to what we had when discussing the position eigenvectors:

$$\int_{-\infty}^{\infty} dp\, u_k^*(p)u_l(p) = \int_{-\infty}^{\infty} dp\, \langle E_k|p\rangle\langle p|E_l\rangle = \langle E_k|E_l\rangle = \delta_{kl}, \tag{2.90}$$

$$\sum_k u_k^*(p)u_k(p') = \sum_k \langle p'|E_k\rangle\langle E_k|p\rangle = \langle p'|p\rangle = \delta(p, p'). \tag{2.91}$$

Finally, I remind you the physical interpretation of the *momentum* wave function,

$$|\psi\rangle \longleftrightarrow \phi(p) = \langle p|\psi\rangle = \frac{1}{\sqrt{\hbar}}\tilde{\psi}(p/\hbar). \tag{2.92}$$

The probability $P_p(a, b)$ for finding the momentum in the range (a, b) is given by

$$P_p(a, b) = \int_a^b dp\, |\phi(p)|^2. \tag{2.93}$$

2.6.3 General formalism

Let us now create a little formalism to summarize these results and to generalize them to other observables which may have a continuous part to their spectrum. All this formalism can be justified using the spectral theory of self-adjoint operators on Hilbert space.

Let $\hat{A}$ be a self-adjoint operator on a Hilbert space $\mathcal{H}$. Its spectrum can consist of a discrete part,

$$\hat{A}|a_n\rangle = a_n|a_n\rangle, \quad n = 1, 2, \ldots, \quad a_n \in \mathbf{R}, \tag{2.94}$$

and a continuous part,

$$\hat{A}|\lambda\rangle = \lambda|\lambda\rangle, \quad \lambda \in U \subset \mathbf{R}. \tag{2.95}$$

The eigenvectors $|a_n\rangle$ are elements of $\mathcal{H}$ and can be chosen to be orthonormal

$$\langle a_m|a_n\rangle = \delta_{mn}. \tag{2.96}$$

The $|\lambda\rangle$ are limits of elements of $\mathcal{H}$ but are not themselves elements of $\mathcal{H}$. The limits of the scalar products of $|\lambda\rangle$ with a (dense) subspace of $\mathcal{H}$ will exist. These "generalized eigenvectors" will be orthonormal "in the delta function sense":

$$\langle\lambda'|\lambda\rangle = \delta(\lambda', \lambda). \tag{2.97}$$

The eigenvectors and generalized eigenvectors together form a basis for the Hilbert space in the sense that for any $|\psi\rangle \in \mathcal{H}$

$$|\psi\rangle = \sum_n \psi_n |a_n\rangle + \int_U d\lambda\, \psi(\lambda)|\lambda\rangle, \tag{2.98}$$

where

$$\langle\psi|\psi\rangle = \sum_n |\psi_n|^2 + \int_U d\lambda\, |\psi(\lambda)|^2. \tag{2.99}$$

Equivalently,

$$\hat{1} = \sum_n |a_n\rangle\langle a_n| + \int_U d\lambda\, |\lambda\rangle\langle\lambda|. \tag{2.100}$$

Given the "spectral decomposition" (2.98), if the state vector of the system is $|\psi\rangle$, the interpretation of ψ_n is that $|\psi_n|^2$ is the probability the system will be found to have the value a_n (in the state $|a_n\rangle$) for the observable represented by $\hat{A}$. If the eigenvalue a_n is *degenerate* – there is more than one linearly independent eigenvector with eigenvalue a_n for $\hat{A}$ – then the total probability for getting a_n upon a measurement of A is obtained by summing $|\psi_n|^2$ over the degenerate subspace. The interpretation of $\psi(\lambda)$ is that $\int_a^b d\lambda\, |\psi(\lambda)|^2$ is the probability for measuring A and getting the value $\lambda \in (a, b)$. If the generalized eigenvalues are degenerate one must sum/integrate this quantity over the (generalized) eigenspace to get the total probability.

Some self-adjoint operators have purely discrete spectrum; the Hamiltonian for the harmonic oscillator is an example. Other operators have a purely continuous spectrum, *e.g.,* the position and momentum operators. Operators also may have both continuous and discrete parts to their spectrum. An example of this would be the Hamiltonian for a hydrogen atom, where the bound states – the usual atomic energy levels – correspond to the discrete spectrum and the scattering (or ionized) states correspond to the continuous spectrum.

2.7 Time evolution

Next we will briefly review the dynamics of a quantum oscillator in each of the Schrödinger and Heisenberg pictures. I will need the Heisenberg picture later since this picture of dynamics is the most immediately accessible in quantum field theory. You should be familiar with the equivalent Schrödinger picture of dynamics, so we can start there.

The time evolution of a state vector in the Schrödinger picture is, of course, defined by the Schrödinger equation:

$$\hat{H}|\psi(t)\rangle = i\hbar\frac{d}{dt}|\psi(t)\rangle. \tag{2.101}$$

Assuming that the Hamiltonian has no explicit time dependence, as is the the case for the harmonic oscillator, the solution to this equation for a given initial state $|\psi(0)\rangle$ can be expressed as

$$|\psi(t)\rangle = e^{-\frac{i}{\hbar}t\hat{H}}|\psi(0)\rangle. \tag{2.102}$$

You may be more familiar with an alternative – but equivalent – form of this solution. If the initial vector is expanded in a basis of energy eigenvectors:

$$|\psi(0)\rangle = \sum_n c_n |E_n\rangle, \tag{2.103}$$

then (2.102) takes the form

$$|\psi(t)\rangle = \sum_n c_n e^{-\frac{i}{\hbar}E_n t}|E_n\rangle. \tag{2.104}$$

In the Schrödinger picture any observable A is represented as a linear operator built from $\hat{x}$ and $\hat{p}$; we write $\hat{A} = \hat{A}(\hat{x}, \hat{p})$. Given the state $|\psi(t)\rangle$ at time t, the expectation value of A at time t, denoted by $\langle A\rangle(t)$, is calculated in the Schrödinger picture by

$$\langle A\rangle(t) = \langle\psi(t)|\hat{A}|\psi(t)\rangle. \tag{2.105}$$

As you may know, all physical predictions of quantum mechanics can be expressed in terms of expectation values of suitable observables, so (2.105) suffices to provide all dynamical information.

The Heisenberg picture of dynamics can be understood as arising from a different organization of terms in the fundamental formula (2.105). We write

$$\langle A\rangle(t) = \langle\psi(t)|\hat{A}|\psi(t)\rangle = \langle\psi(0)|e^{\frac{i}{\hbar}t\hat{H}}\hat{A}e^{-\frac{i}{\hbar}t\hat{H}}|\psi(0)\rangle \equiv \langle\psi(0)|\hat{A}(t)|\psi(0)\rangle, \tag{2.106}$$

where I've defined

$$\hat{A}(t) = e^{\frac{i}{\hbar}t\hat{H}}\hat{A}e^{-\frac{i}{\hbar}t\hat{H}}. \tag{2.107}$$

You can see that using the last equality in (2.106) along with (2.107) amounts to assigning the time dependence to the operator rather than to the state vector. The only thing that matters, physically speaking, is the combination of factors which appears in (2.106), so the organization of the time dependence within this expression is a matter of convenience only. Evidently, we can view the mathematical representation of the time evolution of any observable either as a one parameter family of state vectors with a fixed operator representing the observable, or as a one parameter family of operator representatives of the observable and a fixed state vector. The former organization of the mathematics is the Schrödinger picture and the latter organization is the Heisenberg picture. In the Heisenberg picture the operator $\hat{A}(t)$ represents the observable A at time t. From (2.107) you can see that in general the Heisenberg picture operator(s) for the observable A will be the same as the Schrödinger picture operator for A at a single time, which has been chosen to be $t = 0$.

You might find the Heisenberg picture is more closely aligned with how you learned to think about dynamics in Newtonian mechanics. There we always speak of the time evolution of observables like position and momentum. In the Heisenberg picture of quantum mechanics we do the same thing with the operators representing the observables. In the Heisenberg picture there is no Schrödinger equation – the state vector is the same for all time, and it is determined once and for all by initial conditions. The equations governing time evolution are of the form (2.107). For a system like the harmonic oscillator, all observables are built from $\hat{x}$ and $\hat{p}$, so it suffices to understand

$$\hat{x}(t) = e^{\frac{i}{\hbar}t\hat{H}}\hat{x}e^{-\frac{i}{\hbar}t\hat{H}}, \quad \hat{p}(t) = e^{\frac{i}{\hbar}t\hat{H}}\hat{p}e^{-\frac{i}{\hbar}t\hat{H}}, \tag{2.108}$$

in order to understand the time evolution of the oscillator in the Heisenberg picture. These relations can be expressed as differential equations, just as (2.102) can be expressed as (2.101). You can check

that if the definitions of $x(t)$ and $p(t)$ are differentiated with respect to time then the result is

$$\frac{d}{dt}\hat{x}(t) = \frac{1}{i\hbar}[\hat{x}(t), \hat{H}], \quad \frac{d}{dt}\hat{p}(t) = \frac{1}{i\hbar}[\hat{p}(t), \hat{H}]. \tag{2.109}$$

These are the *Heisenberg equations* of motion. They can be considered the quantum versions of Hamilton's equations. In this analogy, the operators represent the classical position and momentum variables, and the commutator represents the Poisson bracket. To see this a little more clearly, we need to compute the commutators appearing in (2.109). This can be done by noting that for any operator $\hat{A}(t)$ we have

$$[\hat{A}(t), \hat{H}] = [e^{\frac{i}{\hbar}t\hat{H}}\hat{A}(0)e^{-\frac{i}{\hbar}t\hat{H}}, \hat{H}] = e^{\frac{i}{\hbar}t\hat{H}}[\hat{A}(0), \hat{H}]e^{-\frac{i}{\hbar}t\hat{H}} = [\hat{A}(0), \hat{H}](t). \tag{2.110}$$

Applying this to the displacement and momentum for a harmonic oscillator gives

$$[\hat{x}(t), \hat{H}] = \frac{i\hbar}{m}\hat{p}(t), \quad [\hat{p}(t), \hat{H}] = -i\hbar m\omega^2\hat{x}(t), \tag{2.111}$$

so that

$$\frac{d}{dt}\hat{x}(t) = \frac{1}{m}\hat{p}(t), \quad \frac{d}{dt}\hat{p}(t) = -m\omega^2\hat{x}(t), \tag{2.112}$$

which have the same form as Hamilton's equations for a harmonic oscillator. Consequently, these equations are straightforward to solve:

$$\hat{x}(t) = \cos(\omega t)\,\hat{x} + \frac{1}{m\omega}\sin(\omega t)\,\hat{p}, \quad \hat{p}(t) = \cos(\omega t)\,\hat{p} - m\omega\sin(\omega t)\,\hat{x}. \tag{2.113}$$

Here I have used the identification

$$\hat{x}(0) = \hat{x}, \quad \hat{p}(0) = \hat{p}. \tag{2.114}$$

As you can see, the Heisenberg operators evolve in time in the same way as their classical counterparts. You can also see how the Heisenberg operators are the same as the Schrödinger operator at the time $t = 0$.

From the forms of $\hat{x}(t)$ and $\hat{p}(t)$ we easily compute the Heisenberg form, $\hat{a}(t)$, $\hat{a}^\dagger(t)$, of the creation and annihilation operators $\hat{a} = \hat{a}(0)$, $\hat{a}^\dagger = \hat{a}^\dagger(0)$. We have

$$\hat{a}(t) = \sqrt{\frac{m\omega}{2\hbar}}(\hat{x}(t) + \frac{i}{m\omega}\hat{p}(t)) = e^{-i\omega t}\hat{a} \tag{2.115}$$

and

$$\hat{a}^\dagger(t) = \sqrt{\frac{m\omega}{2\hbar}}(\hat{x}(t) - \frac{i}{m\omega}\hat{p}(t)) = e^{i\omega t}\hat{a}^\dagger, \tag{2.116}$$

as you might have anticipated based upon the classical analogs of these formulas.

Finally, it is straightforward to calculate the Heisenberg form of the Hamiltonian:

$$\hat{H}(t) = \frac{1}{2m}\hat{p}^2(t) + \frac{1}{2}m\omega^2\hat{x}^2(t) = \frac{1}{2m}\hat{p}^2(0) + \frac{1}{2}m\omega^2\hat{x}^2(0) = \frac{1}{2m}\hat{p}^2 + \frac{1}{2}m\omega^2\hat{x}^2, \tag{2.117}$$

or

$$\hat{H}(t) = \hbar\omega(\hat{a}^\dagger(t)\hat{a}(t) + \frac{1}{2}\hat{1}) = \hbar\omega(\hat{a}^\dagger(t)\hat{a}(t) + \frac{1}{2}\hat{1}) = \hbar\omega(\hat{a}^\dagger\hat{a} + \frac{1}{2}\hat{1}). \tag{2.118}$$

Evidently, the Heisenberg form of the Hamiltonian operator is the same as its Schrödinger form. This is because

$$\hat{H}(t) = e^{\frac{i}{\hbar}\hat{H}t}\hat{H}e^{-\frac{i}{\hbar}\hat{H}t} = \hat{H} \tag{2.119}$$

owing to the fact that

$$[\hat{H}, f(\hat{H})] = 0 \tag{2.120}$$

for any function f of $\hat{H}$. This result, $\hat{H}(t) = \hat{H}(0)$, is equivalent to conservation of energy for the quantum oscillator.

2.8 Coherent States

There is an important family of states that is useful for making the connection between the classical and quantum oscillator. These are the *coherent states*, mathematically defined as eigenvectors of the lowering operator $\hat{a}$:

$$\hat{a}|z\rangle = z|z\rangle. \tag{2.121}$$

Note that $\hat{a}$ is not symmetric and its eigenvalues are in general complex. Indeed, it can be shown that there is a coherent state associated to any complex number z. In particular, the ground state of the oscillator is a coherent state with $z = 0$. Since $\hat{a}$ is not self-adjoint, the usual issues with generalized eigenvectors and continuous spectrum do not occur; all these eigenvectors are in the Hilbert space and can be normalized in the usual way. However, they are *not* orthogonal for different eigenvalues. The coherent states are *over-complete*, which means that they span the Hilbert space but they are not all linearly independent. It can be shown that any coherent state with $z \neq 0$ will have non-zero probabilities for all energies to occur. (See the Problems.)

The coherent states enjoy a number of important properties. The real and imaginary parts of the eigenvalues yield the expectation values of position and momentum:

$$\langle z|\hat{x}|z\rangle = \sqrt{\frac{2\hbar}{m\omega}}\Re(z), \quad \langle z|\hat{p}|z\rangle = \sqrt{2\hbar m\omega}\Im(z). \tag{2.122}$$

All of these states are *minimum uncertainty states*:

$$(\Delta x)^2 = \langle z|\hat{x}^2|z\rangle - \langle z|\hat{x}|z\rangle^2 = \frac{\hbar}{2m\omega}, \tag{2.123}$$

$$(\Delta p)^2 = \langle z|\hat{p}^2|z\rangle - \langle z|\hat{p}|z\rangle^2 = \frac{\hbar m\omega}{2}, \tag{2.124}$$

$$\Delta x \Delta p = \frac{\hbar}{2}. \tag{2.125}$$

Finally, if the oscillator is in a coherent state defined by z at time $t = 0$, then at time t it is in the coherent state defined by $ze^{-i\omega t}$. This last fact is very easy to see in the Heisenberg picture. This means that the complex eigenvalue z evolves in time in the same way as the classical complex amplitude a, defined in (5.99).

The coherent states can be viewed as states which are "closest to classical" in that the position and momentum have the minimum possible uncertainty and their expectation values evolve according to the classical equations of motion (as they must by Ehrenfest's theorem). For macroscopic values of mass and frequency, these states represent the classical Newtonian behavior of the oscillator to good accuracy.

2.9 Problems

1. A system with 2 degrees of freedom, labeled x and y, has the following Lagrangian:

$$L = \frac{1}{2}m(\dot{x}^2 + \dot{y}^2) - \alpha(x^2 + y^2) - \beta(x - y)^2.$$

   Find the point or points of stable equilibrium. Find the normal modes and characteristic frequencies of oscillation about the equilibria.

2. Prove the result quoted in the text concerning *simultaneous diagonalization of quadratic forms.*

   Hints: (1) The quadratic form M defining the approximate kinetic energy can be used to define a scalar product $(\vec{v}, \vec{w}) \equiv M(\vec{v}, \vec{w}) = v^i w^j M_{ij} = v^T M w$. (2) Every scalar product allows a linear change of basis to an orthonormal basis $\vec{e}_i$ in which $(\vec{e}_i, \vec{e}_j) = \delta_{ij}$. (3) The orthonormal basis is unique up to any change of basis $\vec{e}_i \to O_i^j \vec{e}_j$ where O is an orthogonal matrix, $O^T = O^{-1}$. The potential energy quadratic form is defined by a symmetric array K. Under a change of orthonormal basis defined by O the array changes by $K \to O^T K O = O^{-1} K O$. (4) Use the linear algebra result that any symmetric array, such as K, can be diagonalized via a similarity transformation by an orthogonal matrix.

3. Show that the momentum (2.28) and the Hamiltonian (2.33) are symmetric operators.

4. Show that the time evolution defined by the Schrödinger equation (2.101) preserves the normalization of the state vector:

$$\frac{d}{dt}\langle\psi(t)|\psi(t)\rangle = 0.$$

5. Show that the linear operations defined in (2.51) satisfy the adjoint relation (2.29). (*Hint: It is sufficient to check the relation on a basis.*)

6. Using the probability interpretation of the state vector, show that the expectation value – the statistical mean – of an observable A in a state represented by $|\psi\rangle$ can be calculated by

$$\langle A\rangle = \langle\psi|\hat{A}|\psi\rangle.$$

7. Define the projection operator into the ground state for the harmonic oscillator by

$$\hat{P}_0|\psi\rangle = \langle E_0|\psi\rangle|E_0\rangle.$$

   Show that this is a linear operator. Show that this operator is symmetric. Find the eigenvalues and eigenvectors of this operator. Show that the expectation value of P_0 in the state $|\phi\rangle$ is the probability that an energy measurement in the state represented by $|\phi\rangle$ results in the ground state energy. (In this fashion all probabilities in quantum mechanics can be reduced to computations of expectation values of suitable observables.)

8. The *momentum representation* for quantum mechanics uses the Fourier transform to identify the Hilbert space $\mathcal{H}$ with square integrable functions of momentum. In the momentum representation the position and momentum operators are given by

$$\hat{p}\phi(p) = p\phi(p), \quad \hat{x}\phi(p) = i\hbar\frac{d}{dp}\phi(p).$$

Show that these operators satisfy the commutation relations (2.32). Express the harmonic oscillator Hamiltonian as an operator on momentum wave functions. Using the known spectrum of this Hamiltonian in the position representation, (2.35) and (2.36), deduce the spectrum of the Hamiltonian in the momentum representation. (*Hint:* You do not have to take any Fourier transforms.)

9. Find the ground and first excited states of the oscillator in the momentum representation by taking the Fourier transform of the position representation wave functions.

10. The *statistical uncertainty* in an observable A in a given state is defined to be the standard deviation ΔA of (an arbitrarily large number of) measurements of that observable in an ensemble of systems in the same state. Given a state vector $|\psi\rangle$, the uncertainty can be computed via

$$(\Delta A)^2 = \langle A^2\rangle - \langle A\rangle^2 = \langle\psi|\hat{A}^2|\psi\rangle - \langle\psi|\hat{A}|\psi\rangle^2.$$

For the ground state of the harmonic oscillator calculate ΔH, Δx and Δp.

11. At $t = 0$ a harmonic oscillator is in the state

$$|\psi(0)\rangle = \cos a|E_0\rangle + \sin a|E_1\rangle,$$

where a is some real number. Calculate $\langle x\rangle(t)$, the expectation value of displacement as a function of time.

12. What are the eigenvalues of the kinetic and potential energies for a harmonic oscillator? Why don't they add up to give the discrete spectrum of energies?

13. For a quantum oscillator, consider eigenvectors of the annihilation operator:

$$\hat{a}|z\rangle = z|z\rangle.$$

Show that for any $z \in \mathbf{C}$ there is a normalized eigenvector

$$|z\rangle = e^{-|z|^2/2}e^{z\hat{a}^\dagger}|E_0\rangle. \tag{2.126}$$

Prove that there are no eigenvectors of $\hat{a}^\dagger$. Verify the uncertainties (2.123) and (2.124). Show that the probability $P_n(z)$ for getting n energy quanta in the state $|z\rangle$ is the *Poisson distribution*

$$P_z(n) = e^{-|z|^2}\frac{|z|^{2n}}{n!}. \tag{2.127}$$

Finally, show from the Poisson distribution that in a coherent state defined by z the mean number of energy quanta is

$$\langle n\rangle = \sum_{n=0}^{\infty} nP_z(n) = |z|^2.$$

14. For a quantum oscillator, define new creation and annihilation operators via

$$\hat{b} = \lambda\hat{a} + \nu\hat{a}^\dagger, \quad \hat{b}^\dagger = \lambda\hat{a}^\dagger + \nu\hat{a},$$

where λ and ν are any two real numbers satisfying

$$\lambda^2 - \nu^2 = 1.$$

Show that $[\hat{b}, \hat{b}^\dagger] = \hat{1}$. Show that in the state $|0\rangle$ of vanishing quanta according to $\hat{b}$,

$$\hat{b}|0\rangle = 0,$$

the variances in the displacement and momentum satisfy

$$(\Delta x)^2 = \frac{\hbar}{2m\omega}(\lambda - \nu)^2, \quad (\Delta p)^2 = \frac{\hbar m\omega}{2}(\lambda + \nu)^2$$

so that these states are minimum uncertainty states:

$$(\Delta x)^2(\Delta p)^2 = \frac{\hbar^2}{4}.$$

This property extends to all the coherent states defined by $\hat{b}$; such states are called *squeezed states*.

Chapter 3

Tensor Products and Identical Particles

As we shall see, in a certain sense the simplest incarnation of a quantum field is just an infinite collection of harmonic oscillators. We certainly understand the quantum mechanics of a single oscillator. How do we use this information to understand a system consisting of two or more quantum oscillators? The tensor product is the construction we need. In quantum mechanics we always use the tensor product construction to build a composite system (*e.g.,* many oscillators) from a collection of subsystems (*e.g.,* a single oscillator). This chapter will explain all this. If you are a little rusty on the definition of a vector space, now would be a good time to review it.

3.1 Definition of the tensor product

Let us consider the quantum theory of a system consisting of 2 independent oscillators characterized by masses m_1 and m_2 and frequencies ω_1 and ω_2. (I remind you that the motion of any dynamical system with 2 degrees of freedom near stable equilibrium can be described by two oscillators.) Conceptually at least, it is not hard to describe the quantum structure of the system. For the most part one just considers two copies of the quantum oscillator described in the last chapter. For example, the states of the composite system are obtained by (superpositions of) pairs of vectors, $|\psi_1\rangle$, $|\psi_2\rangle$, coming from two Hilbert spaces, $\mathcal{H}_1$, $\mathcal{H}_2$ of square-integrable functions, $\psi_1(x_1)$, $\psi_2(x_2)$. But how does this exactly fit into the quantum postulates involving a Hilbert space, self-adjoint operators, and all that? This question is answered via the *tensor product* construction.[1]

Consider any two Hilbert spaces $\mathcal{H}_1$ and $\mathcal{H}_2$, with elements denoted by $|\psi\rangle_1$ and $|\phi\rangle_2$, respectively. I will give you two equivalent ways of defining the tensor product. One is simpler than the other, but they both have their advantages and we will definitely want to have both points of view.

First definition

Let $|e_i\rangle$, $i = 1, 2, \ldots, n_1$, and $|f_a\rangle$, $a = 1, 2, \ldots, n_2$, be bases for each of the Hilbert spaces.[2] There is no problem with letting n_1 and/or n_2 become infinite. Consider the set of pairs of basis

[1] The tensor product is also known as the "direct product". It is related to – but is different from – the "Cartesian product".

[2] The ability to find a basis which is in 1-1 correspondence with the set of integers is a slight restriction on the

vectors $|e_i\rangle \otimes |f_a\rangle$, $i = 1, 2, \ldots, n_1$, $a = 1, 2, \ldots, n_2$. Define this to be the basis for $\mathcal{H}_1 \otimes \mathcal{H}_2$ in the sense that every element $|\psi\rangle$ of the tensor product space is defined by a collection of $n_1 \times n_2$ scalars ψ_{ia} via

$$|\psi\rangle = \sum_{i,a} \psi_{ia} |e_i\rangle \otimes |f_a\rangle. \tag{3.1}$$

Addition and scalar multiplication are defined in the usual way for arrays: the arrays are added component-wise and scalar multiplication is performed component-wise.

If you still cling to the column vector way of thinking about vectors, you can do that here. Just view an element $|\psi\rangle \in \mathcal{H}_1 \otimes \mathcal{H}_2$ as a column vector with $n_1 \times n_2$ entries:

$$|\psi\rangle \quad \Longleftrightarrow \begin{pmatrix} c_1 \\ c_2 \\ \cdot \\ \cdot \\ \cdot \\ c_{n_1 \times n_2} \end{pmatrix}, \tag{3.2}$$

where $c_1 = \psi_{11}, c_2 = \psi_{12}, \ldots$.

If you pick a vector $|\alpha\rangle = \sum_i \alpha_i |e_i\rangle$ from $\mathcal{H}_1$ and $|\beta\rangle = \sum_a \beta_a |f_a\rangle$ from $\mathcal{H}_2$ then you can define a *product vector* in $\mathcal{H}_1 \otimes \mathcal{H}_2$ via their components in the chosen basis:

$$\psi_{ia} = \alpha_i \beta_a \quad \Longleftrightarrow \quad |\psi\rangle = \sum_{i,a} \alpha_i \beta_a |e_i\rangle \otimes |f_a\rangle. \tag{3.3}$$

The product vector $|\psi\rangle$ in (3.3) represents the state of the composite system in which subsystem 1 is in state $|\alpha\rangle$ and subsystem 2 is in state $|\beta\rangle$. While not every vector in $\mathcal{H}_1 \otimes \mathcal{H}_2$ is a product vector, every vector is a linear combination of product vectors. Of course, there will exist bases for the tensor product space that do not consist of product vectors. If the vector spaces have respective dimensions n_1 and n_2, then the tensor product space has dimension $n_1 \times n_2$.

The scalar product for $\mathcal{H}_1 \otimes \mathcal{H}_2$ is the obvious one arising from its representation in terms of column vectors. With (3.2) and

$$|\phi\rangle \quad \Longleftrightarrow \begin{pmatrix} d_1 \\ d_2 \\ \cdot \\ \cdot \\ \cdot \\ d_{n_1 \times n_2} \end{pmatrix}, \tag{3.4}$$

we have

$$\langle \phi | \psi \rangle = d_1^* c_1 + d_2^* c_2 + \ldots. \tag{3.5}$$

If $c_1 = \psi_{11}, c_2 = \psi_{12}, \ldots$, then $|c_1|^2$ is the probability that subsystems 1 and 2 are in states $|e_1\rangle$ and $|f_1\rangle$, while $|c_2|^2$ is the probability that subsystems 1 and 2 are in states $|e_1\rangle$ and $|f_2\rangle$, and so on.

---

allowed Hilbert spaces to the class of *separable* Hilbert spaces. It is not necessary that the Hilbert spaces are separable for what will follow, but for convenience I will suppose that they are separable.

I used a choice of basis in each of $\mathcal{H}_1$ and $\mathcal{H}_2$ to define $\mathcal{H}_1 \otimes \mathcal{H}_2$. It is not difficult to show that a different choice of bases will lead to a tensor product vector space which is isomorphic to the original one.

Second Definition

We define a new vector space, the *tensor product of* $\mathcal{H}_1$ *and* $\mathcal{H}_2$, denoted by $\mathcal{H}_1 \otimes \mathcal{H}_2$, as follows. Start with the set of pairs of all possible vectors from each space. With $|\psi\rangle \in \mathcal{H}_1$ and $|\chi\rangle \in \mathcal{H}_2$, a pair is denoted by $|\psi\rangle \otimes |\chi\rangle$. When writing pairs of vectors the first will always come from $\mathcal{H}_1$ and the second will always come from $\mathcal{H}_2$. These pairs of elements are sometimes called "product vectors". Product vectors will be elements of the tensor product vector space – but they do not exhaust its elements. Scalar multiplication on product vectors is defined via the following identifications

$$|\psi\rangle \otimes |\chi\rangle \longrightarrow c\,(|\psi\rangle \otimes |\chi\rangle) \equiv (c|\psi\rangle) \otimes |\chi\rangle \equiv |\psi\rangle \otimes (c|\chi\rangle). \tag{3.6}$$

As you can see, I am simply using the scalar multiplication that is already available in each of the original vector spaces. Notice that scalar multiplication is actually defined by an equivalence class of product vectors. It is this equivalence relation, along with a few more to be introduced momentarily, which distinguish the tensor product from the Cartesian product.

Likewise, we use the vector addition structure on each factor space to make the following identifications on the direct product space:

$$|\psi\rangle \otimes |\chi\rangle + |\phi\rangle \otimes |\chi\rangle \equiv (|\psi\rangle + |\phi\rangle) \otimes |\chi\rangle, \tag{3.7}$$

and

$$|\psi\rangle \otimes |\chi\rangle + |\psi\rangle \otimes |\lambda\rangle \equiv |\psi\rangle \otimes (|\chi\rangle + |\lambda\rangle). \tag{3.8}$$

Of course, the preceding definition does not define addition on the complete space of pairs, only on special pairs – those with a common first or second factor. In general, we view the sum

$$|\psi\rangle \otimes |\chi\rangle + |\phi\rangle \otimes |\sigma\rangle$$

as a new element of the set $\mathcal{H}_1 \otimes \mathcal{H}_2$ (unless of course $|\psi\rangle = |\phi\rangle$ and/or $|\chi\rangle = |\sigma\rangle$, in which case our previous definition applies). Scalar multiplication is then extended to all elements of the set by linearity. So, for example,

$$c\,(|\psi\rangle \otimes |\chi\rangle + |\phi\rangle \otimes |\sigma\rangle) = c|\psi\rangle \otimes |\chi\rangle + c|\phi\rangle \otimes |\sigma\rangle. \tag{3.9}$$

With these definitions and identifications, the tensor product space $\mathcal{H}_1 \otimes \mathcal{H}_2$ is the set of all linear combinations of all pairs of vectors. It is not too hard to show that we get a basis for $\mathcal{H}_1 \otimes \mathcal{H}_2$ by taking the tensor product of the bases for $\mathcal{H}_1$ and $\mathcal{H}_2$. This means that, while not all elements of $\mathcal{H}_1 \otimes \mathcal{H}_2$ are product vectors, every element of $\mathcal{H}_1 \otimes \mathcal{H}_2$ can be expressed as a linear combination of product vectors.

The scalar product for $\mathcal{H}_1 \otimes \mathcal{H}_2$ is induced by the scalar products on the two original Hilbert spaces. The definition goes as follows. Let $|\Psi\rangle, |\Phi\rangle \in \mathcal{H}_1 \otimes \mathcal{H}_2$ be a pair of product vectors:

$$|\Psi\rangle = |\alpha\rangle \otimes |\beta\rangle, \quad |\Phi\rangle = |\gamma\rangle \otimes |\delta\rangle. \tag{3.10}$$

We define

$$\langle\Psi|\Phi\rangle = \langle\alpha|\gamma\rangle\langle\beta|\delta\rangle. \tag{3.11}$$

We extend this definition to linear combinations of product states by linearity in $|\Phi\rangle$ and anti-linearity in $\langle\Psi|$. For example,

$$|\Psi\rangle = a|\alpha\rangle \otimes |\beta\rangle + b|\chi\rangle \otimes |\psi\rangle, \quad |\Phi\rangle = c|\mu\rangle \otimes |\rho\rangle + d|\sigma\rangle \otimes |\tau\rangle, \tag{3.12}$$

$$\langle\Psi|\Phi\rangle = a^*c\langle\alpha|\mu\rangle\langle\beta|\rho\rangle + a^*d\langle\alpha|\sigma\rangle\langle\beta|\tau\rangle + b^*c\langle\chi|\mu\rangle\langle\psi|\rho\rangle + b^*d\langle\chi|\sigma\rangle\langle\psi|\tau\rangle. \tag{3.13}$$

Since every element of $\mathcal{H}_1 \otimes \mathcal{H}_2$ is a linear combination of product states this defines the scalar product on all vectors in $\mathcal{H}_1 \otimes \mathcal{H}_2$.

The first definition of the tensor product is relatively simple, and shows how to calculate using a basis and components. The second definition is more elegant since it does not require the introduction of a basis, and it shows how to calculate on the vector space in general, without explicitly introducing arrays of components. The two definitions can be shown to be equivalent in the sense that the vector space of Definition 1 is naturally isomorphic to the one coming from Definition 2. One does this by showing that the product basis of Definition 1 is also a product basis for the vector space of Definition 2.

Elements of $\mathcal{H}_1 \otimes \mathcal{H}_2$ are often referred to as "tensors of rank 2". This terminology is correct, but perhaps confusing in the quantum mechanical context. For our purposes, the key idea is that the tensors of rank 2 form a vector space – they can be added and multiplied by scalars – and it is this vector space structure which is paramount in quantum mechanics. So I shall continue to refer to elements of $\mathcal{H}_1 \otimes \mathcal{H}_2$ as "vectors", even though it is correct to view them as tensors.

Let us use square-integrable functions, *e.g.,* $\psi(x_1)$ and $\phi(x_2)$, to represent the vectors $|\psi\rangle \in \mathcal{H}_1$ and $|\phi\rangle \in \mathcal{H}_2$, and see how to represent states in $\mathcal{H}_1 \otimes \mathcal{H}_2$. Let $|x_1\rangle$, $|x_2\rangle$ be the generalized position eigenvectors for each oscillator. As we will justify further in the next section, we can view $|x_1, x_2\rangle \equiv |x_1\rangle \otimes |x_2\rangle$ as a (generalized) basis for $\mathcal{H}_1 \otimes \mathcal{H}_2$. We have

$$\psi(x_1) = \langle x_1|\psi\rangle, \quad \phi(x_2) = \langle x_2|\phi\rangle, \tag{3.14}$$

so that the function representing the state $|\psi\rangle \otimes |\phi\rangle$ is given by

$$|\psi\rangle \otimes |\phi\rangle \quad \longleftrightarrow \quad \langle x_1, x_2|(|\psi\rangle \otimes |\phi\rangle) = \langle x_1|\psi\rangle\langle x_2|\phi\rangle = \psi(x_1)\phi(x_2). \tag{3.15}$$

The most general wave function on $\mathcal{H}_1 \otimes \mathcal{H}_2$ is then a linear combination of such functions. It can be shown that the set of all such linear combinations (when completed in the Hilbert space norm) is isomorphic to the Hilbert space of square-integrable functions of two variables. This means that for a general $|\Lambda\rangle \in \mathcal{H}_1 \otimes \mathcal{H}_2$ we have

$$|\Lambda\rangle \quad \longleftrightarrow \quad \Lambda(x_1, x_2) = \langle x_1, x_2|\Lambda\rangle, \quad \text{where } \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} dx_1\, dx_2\, |\Lambda(x_1, x_2)|^2 < \infty. \tag{3.16}$$

Thus the Hilbert space of states for two oscillators is equivalent to the space of square-integrable functions on $\mathbf{R}^2$ with the standard inner product:

$$\langle\Lambda|\Omega\rangle = \int_{\mathbf{R}^2} d^2x\, \Lambda^*(x_1, x_2)\Omega(x_1, x_2). \tag{3.17}$$

The interpretation of the wave function $\Lambda(x_1, x_2)$ is that $|\Lambda(x_1, x_2)|^2$ is the probability density for finding particle one at x_1 and particle 2 at x_2.

If you are comfortable with the tensor product of two vector spaces, I think you can easily see how to generalize the preceding discussion to describe a system consisting of N subsystems. You simply repeat all the previous constructions, but use the space $\mathcal{H}_1 \otimes \mathcal{H}_2 \otimes \cdots \otimes \mathcal{H}_N$.

3.2 Observables and the tensor product

Having constructed the vector space of states $\mathcal{H}_1 \otimes \mathcal{H}_2$ of the composite system from the subsystem spaces $\mathcal{H}_1$ and $\mathcal{H}_2$, we now need to see how to construct operator representatives of observables for the composite system. I will show you how it goes in the context of two oscillators, but the results are quite general.

Each oscillator has its respective observables, represented by linear operators constructed from $(\hat{x}_1, \hat{p}_1)$ and $(\hat{x}_2, \hat{p}_2)$. Our first task is to see how these subsystem observables are represented on the vector space of states for the composite system. Given a basis $|i\rangle$, $i = 1, 2, \ldots$, for a single oscillator (*e.g.*, the basis of energy eigenvectors) any state vector $|\Psi\rangle \in \mathcal{H}_1 \otimes \mathcal{H}_2$ can be expanded in a product basis:

$$|\Psi\rangle = \sum_{i,j} \Psi_{ij} \, |i\rangle \otimes |j\rangle, \quad |i\rangle \in \mathcal{H}_1, \quad |j\rangle \in \mathcal{H}_2 \tag{3.18}$$

Let $\hat{A}_1$ be an operator defined on $\mathcal{H}_1$. We extend its definition to $|\Psi\rangle \in \mathcal{H}_1 \otimes \mathcal{H}_2$ via

$$\hat{A}|\Psi\rangle = \sum_{i,j} \Psi_{ij} \, (\hat{A}_1|i\rangle) \otimes |j\rangle. \tag{3.19}$$

Similarly if $\hat{B}_2$ is an operator defined on $\mathcal{H}_2$ we extend its definition to $|\Psi\rangle \in \mathcal{H}_1 \otimes \mathcal{H}_2$ via

$$\hat{B}|\Psi\rangle = \sum_{i,j} \Psi_{ij} \, |i\rangle \otimes (\hat{B}_2|j\rangle). \tag{3.20}$$

These extensions of $\hat{A}_1$ and $\hat{B}_2$ to $|\Psi\rangle \in \mathcal{H}_1 \otimes \mathcal{H}_2$ use the pretty obvious notation

$$\hat{A} = \hat{A}_1 \otimes \hat{1}, \quad \hat{B} = \hat{1} \otimes \hat{B}_2. \tag{3.21}$$

With this definition you can see that the extension of a pair of observables $\hat{A}_1$ and $\hat{B}_2$ for systems 1 and 2 to the composite system's vector space $\mathcal{H}_1 \otimes \mathcal{H}_2$ will have a product denoted by $\hat{A}\hat{B} = \hat{A}_1 \otimes \hat{B}_2$, where

$$\hat{A}_1 \otimes \hat{B}_2|\psi\rangle = \sum_{i,j} \Psi_{ij} \, (\hat{A}_1|i\rangle) \otimes (\hat{B}_2|j\rangle). \tag{3.22}$$

The two observables A and B will always be compatible, that is, their operator representatives will commute:

$$[\hat{A}, \hat{B}] = (\hat{A}_1 \otimes \hat{1})(\hat{1} \otimes \hat{B}_2) - (\hat{1} \otimes \hat{B}_2)(\hat{A}_1 \otimes \hat{1}) = \hat{A}_1 \otimes \hat{B}_2 - \hat{A}_1 \otimes \hat{B}_2 = 0. \tag{3.23}$$

It follows that the positions, momenta, energies *etc.* for each oscillator will be compatible and one may use any two bases of eigenvectors to span the Hilbert space of the composite system. For example one can use position eigenvectors: $|x_1\rangle \otimes |x_2\rangle$, or position and momentum eigenvectors $|x_1\rangle \otimes |p_2\rangle$ or energy eigenvectors $|E_1\rangle \otimes |E_2\rangle$, and so forth. Such bases represent states in which the respective observables for each particle are known with statistical certainty.

Besides the usual observables pertaining to subsystems, new observables which necessarily refer to the composite system are also available. For example, if the two subsystems do not interact, *i.e.,* they do not exchange energy, the total energy of the system is defined as

$$\hat{H} = \hat{H}_1 \otimes \hat{1} + \hat{1} \otimes \hat{H}_2. \tag{3.24}$$

Eigenvectors for this operator are given by the product of energy eigenvectors for the subsystem:

$$\begin{aligned}\hat{H}|E_m\rangle \otimes |E_n\rangle &= (\hat{H}_1|E_m\rangle) \otimes |E_n\rangle + |E_m\rangle \otimes (H_2|E_n\rangle) \\ &= E_m|E_m\rangle \otimes |E_n\rangle + E_n|E_m\rangle \otimes |E_n\rangle \\ &= (E_m + E_n)|E_m\rangle \otimes |E_n\rangle. \end{aligned} \tag{3.25}$$

The total energy eigenvalue is just the sum of the individual eigenvalues in this non-interacting scenario. Physically, these product vectors are states in which the individual energies and the total energy are known with certainty. A basis of such states is available because the three relevant operators commute.

I think you can easily see how to generalize the preceding discussion for a system consisting of N subsystems. You simply repeat all the previous constructions, but use the space $\mathcal{H}_1 \otimes \mathcal{H}_2 \otimes \cdots \otimes \mathcal{H}_N$. Observables pertaining to, say, subsystem k would take the form

$$\hat{1} \otimes \hat{1} \otimes \cdots \hat{1} \otimes \hat{A}_k \otimes \hat{1} \cdots \otimes \hat{1}. \tag{3.26}$$

This Hilbert space and operators on it is the setting for the quantum theory of a dynamical system near stable equilibrium, where each of the component Hilbert spaces is that of a quantum oscillator, *i.e.,* square-integrable functions of one variable. In this way one gets the quantum theory of normal modes of vibration for a mechanical system. As we shall see in the next chapter, if we consider a limit in which N become arbitrarily large we get one way of thinking about a quantum field.

3.3 Symmetric and antisymmetric tensors. Identical particles.

Recall that *identical particles* are those which are intrinsically alike in the sense that they all share certain observable characteristics – their defining properties – although they can differ in other observable regards. So, for example, all electrons have the same mass, electric charge, and spin, although they can be in different locations, have different velocities, and have different spin states. The defining properties are what distinguish different kinds of particles, while the other observable properties distinguish different states of the particle. Consider a system of such identical particles. If the roles of any two particles in defining the state of the system are interchanged, the state of the system should be unchanged. For example, if the state of the system is determined by specifying that one electron is on Earth and another electron is on Mars then, if we swap the two electrons, we still have one electron on Earth and one on Mars and the state is unchanged. We can elevate this conclusion about the states of systems of identical particles to a new postulate of quantum mechanics. This postulate involves the symmetric and antisymmetric tensor product, which I will now explain.

By definition, identical particles will have identical Hilbert spaces of state vectors. Let us consider two identical particles and their respective Hilbert spaces $\mathcal{H}_1$ and $\mathcal{H}_2$. The system of two particles is described by the Hilbert space $\mathcal{H}_1 \otimes \mathcal{H}_2$, but we must take account of the indistinguishability of the particles. Consider a product vector $|\alpha\rangle \otimes |\beta\rangle \in \mathcal{H}_1 \otimes \mathcal{H}_2$. This represents a state in which one of the particles (labeled "1") is in the single particle state $|\alpha\rangle$ and the other particle (labeled "2") is in the single particle state $|\beta\rangle$. We are interested in what happens to the state when the roles of particle 1 and particle 2 are interchanged. To this end, we can decompose the product

vector into its *symmetric* and *antisymmetric* parts:

$$\begin{aligned}|\alpha\rangle \otimes |\beta\rangle &= \frac{1}{2}(|\alpha\rangle \otimes |\beta\rangle + |\beta\rangle \otimes |\alpha\rangle) + \frac{1}{2}(|\alpha\rangle \otimes |\beta\rangle - |\beta\rangle \otimes |\alpha\rangle) \\ &\equiv |\alpha,\beta\rangle_s + |\alpha,\beta\rangle_a. \end{aligned} \tag{3.27}$$

The symmetric vectors do not change when the roles of the particles are interchanged. The antisymmetric vectors change by a sign when the roles of the particles are interchanged. A sign change is a phase factor; as mentioned back in the postulates any two vectors which differ by a phase factor define the same state. Consequently the symmetric part of the product vector or the anti-symmetric part of the state vector could serve as a state for two identical particles with one particle in a state $|\alpha\rangle$ and one particle in the state $|\beta\rangle$. The decomposition (3.27) can be extended to each term in the expansion of a vector in a product basis, so we can partition $\mathcal{H}_1 \otimes \mathcal{H}_2$ into its symmetric part $[\mathcal{H}_1 \otimes \mathcal{H}_2]_s$ and its antisymmetric part $[\mathcal{H}_1 \otimes \mathcal{H}_2]_a$. Each of these subspaces is itself a Hilbert space.

For composite systems consisting of more than two particles, one can still consider the subspace of vectors which are unchanged under particle permutation ("symmetric"), or change sign under exchange of any two particles ("antisymmetric"). Unlike the case with just two particles, these two subspaces do not exhaust the whole vector space – there are states with intermediate "Young tableaux symmetries". So far, it appears that nature does not take advantage of this latter type of state vector symmetry.

We are now ready to add a new quantum mechanics postulate: *The states for identical quantum systems are given by either the symmetric or antisymmetric subspace of the tensor product.*

You may recall that in statistical thermodynamics identical particles described by symmetric states obey Bose-Einstein statistics and are called "bosons" while those described by the antisymmetric states obey Fermi-Dirac statistics and are called "fermions". It is a theorem of relativistic quantum field theory that integral spin particles are bosons and half-integral spin particles are fermions. We will explore this fact just a little when we finally get around to constructing photons, which have spin-1. Whether a particle is a boson or a fermion is one of those intrinsic features that makes the particle what it is.[3]

The number of states available to a system of particles depends strongly on whether or not they are identical, and, if they are identical, whether or not they are bosons or fermions. Given two single particle states $|\alpha\rangle$, $|\beta\rangle$, there is a two-dimensional space of physically inequivalent states for a system consisting of 2 distinguishable particles:[4]

$$|\Psi\rangle_{\text{distinguishable}} = a|\alpha\rangle \otimes |\beta\rangle + b|\beta\rangle \otimes |\alpha\rangle, \quad |a|^2 + |b|^2 = 1. \tag{3.28}$$

In this family of states $|a|^2$ is the probability for finding particle 1 in state $|\alpha\rangle$ and particle 2 in state $|\beta\rangle$, while $|b|^2$ is the probability for finding particle 1 in state $|\beta\rangle$ and particle 2 in state $|\alpha\rangle$. In light of the new quantum postulate for identical particles, there is only one way to construct a state of identical particles from the given 1-particle states:

$$|\Psi\rangle_{\text{identical}} = \frac{1}{\sqrt{2}}\left(|\alpha\rangle \otimes |\beta\rangle \pm |\beta\rangle \otimes |\alpha\rangle\right), \tag{3.29}$$

[3]Here it might be good to mention that in supersymmetric theories the distinction between bosons and fermions is not so clear-cut!

[4]There are two complex constants, so *a priori* there four real parameters. Normalization fixes one real parameter, and the equivalence of any two vectors if they differ by a phase removes another real parameter.

where the plus/minus sign goes with bosons/fermions. This is a state in which one particle is in the state $|\alpha\rangle$ and one particle is in the state $|\beta\rangle$. Moreover, if the particles are identical fermions then they cannot be in the same state, $|\alpha\rangle \neq |\beta\rangle$ – this is the quantum mechanical implementation of the Pauli exclusion principle.

This new quantum postulate for systems composed of identical subsystems puts limitations on the allowed observables since not every operator restricts to act on the symmetric or antisymmetric subspace. Let us consider a simple example. For two distinct oscillators with frequencies ω_1 and ω_2, each in an energy eigenstate, one has a state described by the vector

$$|\Psi\rangle = |n\rangle \otimes |j\rangle, \tag{3.30}$$

in which the first oscillator has energy $(n+\frac{1}{2})\hbar\omega_1$ with statistical certainty and the second oscillator has energy $(j+\frac{1}{2})\hbar\omega_2$ with certainty. If we want to model this situation for identical oscillators we should set $m_1 = m_2 \equiv m$ and $\omega_1 = \omega_2 = \omega$ and then restrict to symmetric or antisymmetric states:

$$|\Psi\rangle = \frac{1}{\sqrt{2}}\Big\{|n\rangle \otimes |j\rangle + |j\rangle \otimes |n\rangle\Big\}, \quad \text{bosons,} \tag{3.31}$$

$$|\Psi\rangle = \frac{1}{\sqrt{2}}\Big\{|n\rangle \otimes |j\rangle - |j\rangle \otimes |n\rangle\Big\}, \quad \text{fermions.} \tag{3.32}$$

The meaning of the state now is that one of the oscillators has energy $(n+\frac{1}{2})\hbar\omega$ and the other has energy $(j+\frac{1}{2})\hbar\omega$. If the oscillators are identical, by definition there is no way to tell which is which and this sort of description of, say, an energy eigenstate is the best one can do. Indeed, there can be no observables which pertain to this or that particle individually since they are all identical. For example, the operator representing energy of oscillator 1, $\hat{H} \otimes \hat{1}$ is no longer well-defined since it does not map the symmetric or antisymmetric subspace into itself:

$$(\hat{H} \otimes \hat{1})|\Psi\rangle = \frac{1}{\sqrt{2}}\Big\{(n+\frac{1}{2})\hbar\omega|n\rangle \otimes |j\rangle \pm (j+\frac{1}{2})\hbar\omega|j\rangle \otimes |n\rangle\Big\}. \tag{3.33}$$

To interchange the roles of the oscillators in (3.33) you replace $|n\rangle \otimes |j\rangle \leftrightarrow |j\rangle \otimes |n\rangle$ and you will see that the vector on the right hand side does not stay the same (up to a sign) when $n \neq j$.

There are, of course, observables which can be used to describe a system comprised of identical subsystems. For example, the total energy of the system,

$$\hat{H}_{\text{total}} = \hat{H} \otimes \hat{1} + \hat{1} \otimes \hat{H}, \tag{3.34}$$

is a well-defined operator on the symmetric and antisymmetric subspaces. You can check that the product state we just considered is an eigenvector:

$$\hat{H}_{\text{total}}|\Psi\rangle = \Big[(n+\frac{1}{2})\hbar\omega + (j+\frac{1}{2})\hbar\omega\Big]\,\frac{1}{\sqrt{2}}\Big\{|n\rangle \otimes |j\rangle \pm |j\rangle \otimes |n\rangle\Big\} = (n+j+1)\hbar\omega|\Psi\rangle. \tag{3.35}$$

Evidently this state is one in which the energy is $(n+j+1)\hbar\omega$ with statistical certainty. Since these (anti-)symmetrized product vectors span the Hilbert space for this system of identical particles the operator $\hat{H}_{\text{total}}$ is well-defined. It is not too hard to show that an operator is well-defined on the symmetric or anti-symmetric subspace of states if and only if it is invariant under interchange of the roles of the particles.

3.4 Symmetrization and anti-symmetrization for any number of particles

It is worth briefly describing how one builds the symmetric and antisymmetric state vectors for systems consisting of more than two (identical) subsystems. To begin, let us consider a product state of 3 harmonic oscillators; let the single-particle states be represented energy eigenvectors:

$$|\Psi\rangle = |n_1\rangle \otimes |n_2\rangle \otimes |n_3\rangle. \tag{3.36}$$

This state could apply to a system of of 3 distinguishable oscillators each of which has, with probability one, energy $(n_1 + \frac{1}{2})\hbar\omega_1$, $(n_2 + \frac{1}{2})\hbar\omega_2$, $(n_1 + \frac{1}{2})\hbar\omega_3$. To make the oscillators "identical", set $\omega_1 = \omega_2 = \omega_3$ and $m_1 = m_2 = m_3$. The totally symmetric and antisymmetric parts of this state are given by

$$\begin{aligned}|\Psi\rangle_s =& \frac{1}{\sqrt{3!}}\Big\{|n_1\rangle \otimes |n_2\rangle \otimes |n_3\rangle + |n_2\rangle \otimes |n_3\rangle \otimes |n_1\rangle + |n_3\rangle \otimes |n_1\rangle \otimes |n_2\rangle \\ &+ \big(|n_1\rangle \otimes |n_3\rangle \otimes |n_2\rangle + |n_2\rangle \otimes |n_1\rangle \otimes |n_3\rangle + |n_3\rangle \otimes |n_2\rangle \otimes |n_1\rangle\big)\Big\}\end{aligned} \tag{3.37}$$

$$\begin{aligned}|\Psi\rangle_a =& \frac{1}{\sqrt{3!}}\Big\{|n_1\rangle \otimes |n_2\rangle \otimes |n_3\rangle + |n_2\rangle \otimes |n_3\rangle \otimes |n_1\rangle + |n_3\rangle \otimes |n_1\rangle \otimes |n_2\rangle \\ &- \big(|n_1\rangle \otimes |n_3\rangle \otimes |n_2\rangle + |n_2\rangle \otimes |n_1\rangle \otimes |n_3\rangle + |n_3\rangle \otimes |n_2\rangle \otimes |n_1\rangle\big)\Big\}\end{aligned} \tag{3.38}$$

In the symmetric case I summed over all permutations – the first three terms involve the "even" permutations, which involve an even number of interchanges, and the second three terms involve the "odd" permutations, which involve an odd number of interchanges. In the antisymmetric case I subtracted the sum over odd permutations from the sum over even permutations. These states would describe energy eigenstates of three identical bosons and fermions, respectively.

Given three distinct 1-particle energy eigenstates, there are $3! = 6$ possible ways to assign particles to those states. So, in the case where all three particles were somehow distinguishable there would be 6 product states for the composite system which I could construct from the three 1-particle energy eigenstates. If the particles are indistinguishable there is only one state available!

In general, the symmetric and antisymmetric parts of a product state of p particles are given by

$$|\Psi\rangle = |n_1\rangle \otimes |n_2\rangle \otimes \cdots \otimes |n_p\rangle, \tag{3.39}$$

$$|\Psi\rangle_s = \frac{1}{\sqrt{p!}}\left\{\sum_{\substack{\text{even}\\ \text{permutations}}} |n_1\rangle \otimes |n_2\rangle \otimes \cdots \otimes |n_p\rangle + \sum_{\substack{\text{odd}\\ \text{permutations}}} |n_1\rangle \otimes |n_2\rangle \otimes \cdots \otimes |n_p\rangle\right\}, \tag{3.40}$$

$$|\Psi\rangle_a = \frac{1}{\sqrt{p!}}\left\{\sum_{\substack{\text{even}\\ \text{permutations}}} |n_1\rangle \otimes |n_2\rangle \otimes \cdots \otimes |n_p\rangle - \sum_{\substack{\text{odd}\\ \text{permutations}}} |n_1\rangle \otimes |n_2\rangle \otimes \cdots \otimes |n_p\rangle\right\}, \tag{3.41}$$

The symmetric or anti-symmetric subspaces of the Hilbert space of states are obtained by considering linear combinations of states (3.40) or (3.41), respectively. While I have constructed the symmetric and anti-symmetric states via harmonic oscillator energy eigenvectors, I think you can see how to generalize this to any 1-particle states.

3.5 Problems

1. Construct the Schrödinger equation for two harmonic oscillators characterized by masses m_1, m_2 and frequencies ω_1, ω_2. Find/describe its general solution. (Notice that the wave function defines waves in $\mathbf{R}^2$ rather than two waves in $\mathbf{R}^1$.)

2. Consider two identical oscillators, one in a state represented by the position wave function ψ_1, one in a state represented by the wave function ψ_2. Construct the position wave function of the composite system in the case of distinguishable particles, in the case of bosons, and in the case of fermions. For each case compute the probability density that an oscillator is at x_1 and the other oscillator is at x_2. Show that in the case of bosons the probability is enhanced for the particles to be near each other compared to the distinguishable case, while the probability is reduced in the case of fermions. Finally, show that this effect becomes negligible if the particles are sufficiently localized away from each other in the sense that the overlap of the supports of the two 1-particle wave functions is negligible.

3. Consider N fermionic oscillators in 1-particle states, represented by wave functions $\psi_1, \psi_2, \ldots, \psi_N$. Show that the wave function of the composite system can be represented by the *Slater determinant*

$$\Psi(x_1, x_2, \ldots, x_N) = \frac{1}{\sqrt{N!}} \det \begin{pmatrix} \psi_1(x_1) & \psi_2(x_1) & \ldots & \psi_N(x_1) \\ \psi_1(x_2) & \psi_2(x_2) & \ldots & \psi_N(x_2) \\ . & . & . & . \\ . & . & . & . \\ . & . & . & . \\ \psi_1(x_N) & \psi_2(x_N) & \ldots & \psi_N(x_N) \end{pmatrix}.$$

   Under what circumstances will this wave function be identically zero? (If your muscles are not strong enough, you can restrict your answer to the case $N = 3$.)

4. Show that a well-defined operator for a system of identical bosons or fermions must be invariant under any interchange of two particles.

5. A particle has a two-dimensional Hilbert space with orthonormal basis $|1\rangle$, $|2\rangle$. Consider a system consisting of two such particles, assumed distinguishable. Set up the Hilbert space of the composite system using both definitions of the tensor product and indicate how vectors in each version correspond. Show that in each definition the product basis induced by $|1\rangle$, $|2\rangle$ is orthonormal.

6. Show that the Hamiltonian (3.34) is a well-defined operator on the symmetric and anti-symmetric subspaces of the Hilbert space. This means you should show that $\hat{H}_{\text{total}}$ maps symmetric states to symmetric states and anti-symmetric states to anti-symmetric states.

7. A (non-relativistic) particle with spin-$\frac{1}{2}$ is characterized by a trio of linear operators on the 2-d Hilbert space $\mathbf{C}^2$ representing components of the *spin vector* $\vec{s}$:

$$\hat{s}_x = \frac{\hbar}{2}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \hat{s}_y = \frac{\hbar}{2}\begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \hat{s}_z = \frac{\hbar}{2}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

   Consider two distinguishable spin-$\frac{1}{2}$ particles. Construct the individual spin operators, $\vec{S}_1$ and $\vec{S}_2$ as matrices on the 4-dimensional space $\mathbf{C}^4 = \mathbf{C}^2 \otimes \mathbf{C}^2$ in the product basis of eigenvectors

of S_{1z} and S_{2z}. Construct the matrix representing the *total spin*, $\vec{S} = \vec{S}_1 + \vec{S}_2$ again in the product basis of eigenvectors of S_{1z} and S_{2z}. How should $\vec{S}$ be represented if the particles are identical?

8. For a system of two identical harmonic oscillators, show that while the individual momenta are not well-defined observables the total momentum is well-defined.

9. Consider two particles, for simplicity moving in one dimension. Construct the operators representing the center of mass $\hat{X}$, the total momentum $\hat{P}$, the relative position $\hat{x}$, and the relative momentum $\hat{p}$, all in terms of the individual position and momentum operators $(\hat{x}_1, \hat{x}_2)$ and $(\hat{p}_1, \hat{p}_2)$. Construct a (generalized) eigenfunction of $\hat{x}$ and $\hat{P}$ with eigenvalues a and 0 respectively, corresponding to the particles having zero total momentum and a statistically certain separation a.

   (*Why this is an interesting problem:* In this state, suppose that particle 2's momentum is measured to have the value p_2. Then particle 1 has momentum $-p_2$ with statistical certainty and its position is completely uncertain. Suppose instead that particle 2's position is measured and found to have the value x_2. Then particle 1's position is known with certainty to have the value $x_2 - a$ and its momentum is completely uncertain. Thus measurements made on particle 2 appear to affect the outcome of measurements made on particle 1. This is the essence of the Einstein-Podolsky-Rosen paradox [4].)

10. What is the ground state and ground state energy of a system consisting of three non-interacting oscillators if they are all distinguishable. What if the particles are identical bosons? How about identical fermions?

11. You know that the states of motion of a particle in three dimensions can be represented by a wave function $\psi(\vec{x})$. You also know that the spin states of a spin-$\frac{1}{2}$ system can be represented by a column vector χ with two complex entries. Using tensor product constructions, explain how states of a spin-$\frac{1}{2}$ particle moving in 3-d should be represented.

12. For any physical system, the operators representing angular momentum are a trio $\vec{J} = (\hat{J}_x, \hat{J}_y, \hat{J}_z)$ satisfying
$$[\hat{J}_a, \hat{J}_b] = i\hbar\epsilon_{abc}\hat{J}_c, \tag{3.42}$$
where ϵ_{abc} is the Levi-Civita permutation symbol. Define the length-squared of angular momentum by the operator
$$\hat{J}^2 = \hat{J}_x^2 + \hat{J}_y^2 + \hat{J}_z^2.$$
Using (3.42), show that $[J_a, J^2] = 0$ for any $a = x, y, z$. Consider a composite system consisting of two systems with angular momenta $\vec{J}_1$ and $\vec{J}_2$ acting on Hilbert spaces $\mathcal{H}_1$ and $\mathcal{H}_2$. Define the total angular momentum of the system in the usual way by
$$\vec{J} = \vec{J}_1 \otimes \hat{1} + \hat{1} \otimes \vec{J}_2.$$
Show that the total angular momentum $\vec{J}$ satisfies (3.42). Show that
$$[\hat{J}^2, J_{1a}] \neq 0, \quad [\hat{J}^2, J_{2a}] \neq 0,$$
so that states of definite total angular momentum will not necessarily be states of definite individual angular momentum.

Chapter 4

Fock Space

In quantum mechanics one considers systems consisting of a fixed number of particles. If there are N particles, the N-fold tensor product of the 1-particle Hilbert spaces suffices to define the states. Particle observables are represented as operators on the tensor product. In quantum field theory we consider systems where the number of particles in the system is variable. How to build a Hilbert space of states and operator observables which allows this? Fock space.

4.1 Definitions

Fock space $\mathcal{F}$ is canonically defined from any given Hilbert space $\mathcal{H}$ using direct sums and tensor products.(I will use the Hilbert space $\mathcal{H}$ of square integrable functions on the real line, suitable for the harmonic oscillator, as a running example.) In this context $\mathcal{H}$ is called the *1-particle Hilbert space*; it represents the set of states of a single particle, *e.g.,* an oscillator. As we have seen, the space $\mathcal{H} \otimes \mathcal{H}$ can be viewed either as the set of states for a single particle in higher dimensions, or as the set of states of two particles. Let's use the latter interpretation, and let us assume the particles are identical so we only work with the symmetric or antisymmetric subspace according to whether the particles are bosons or fermions. Similarly we can view $\mathcal{H} \otimes \mathcal{H} \otimes \mathcal{H}$ as the space of states of 3 particles; if the particles are identical we again take the symmetric or antisymmetric subspace. As you can imagine, we can keep on building Hilbert spaces of states for any number of identical particles by taking the appropriate number of tensor products and then taking the appropriate symmetric or antisymmetric subspace. Fock space is defined as the *direct sum* of the set of complex numbers $\mathbf{C}$ (viewed as a 1-d Hilbert space) with the one particle Hilbert space, with the two particle Hilbert space, with the three particle Hilbert space, and so forth.[1] Before seeing how that all works in detail, let me make sure you are comfortable with the notion of the direct sum of two vector spaces.

Given two vector spaces V_1 and V_2 of dimension n_1 and n_2, the *direct sum* of V_1 and V_2 is the vector space $V = V_1 \oplus V_2$ of dimension $n_1 + n_2$ obtained by taking the Cartesian product of the two vector spaces. An element $|w\rangle \in V_1 \oplus V_2$ is then an ordered pair $(|w_1\rangle, |w_2\rangle)$ where $|w_1\rangle \in V_1$

[1] Technically, the Fock space is defined as the completion of this direct sum of tensor products using the norm induced by the scalar product, which will be defined shortly.

and $|w_2\rangle \in V_2$. If you like to think in terms of column vectors, we could write

$$|w_1\rangle = \begin{pmatrix} a_1 \\ a_2 \\ . \\ . \\ . \\ a_{n_1} \end{pmatrix}, \quad |w_2\rangle = \begin{pmatrix} b_1 \\ b_2 \\ . \\ . \\ . \\ b_{n_2} \end{pmatrix}, \quad |w\rangle = \begin{pmatrix} a_1 \\ a_2 \\ . \\ . \\ . \\ a_{n_1} \\ b_1 \\ b_2 \\ . \\ . \\ . \\ b_{n_2} \end{pmatrix}, \tag{4.1}$$

As a familiar example, the usual x-y-z vector space $\mathbf{R}^3$ can be written as a direct sum in a couple of ways: $\mathbf{R}^3 = \mathbf{R} \oplus \mathbf{R}^2 = \mathbf{R} \oplus \mathbf{R} \oplus \mathbf{R}$.

Addition and scalar multiplication on $V = V_1 \oplus V_2$ are defined by

$$|w\rangle + |u\rangle = (|w_1\rangle + |u_1\rangle, |w_2\rangle + |u_2\rangle), \quad c|w\rangle = (c|w_1\rangle, c|w_2\rangle), \tag{4.2}$$

this endows $V_1 \oplus V_2$ with the structure of a (complex) vector space. The scalar product used to provide the Hilbert space structure is defined via

$$\langle w|u\rangle = \langle w_1|u_1\rangle + \langle w_2|u_2\rangle. \tag{4.3}$$

You can check that all this vector space structure is just the usual one for the set of column vectors (4.1). With the direct sum understood, we can now construct the Fock space.

Let us denote the Hilbert space of complex numbers by $\mathbf{C} \equiv \mathcal{H}_0$, the one particle Hilbert space can be denoted by $\mathcal{H} \equiv \mathcal{H}_1$, the two particle Hilbert space can be denoted by $\mathcal{H}_2 \equiv \mathcal{H} \otimes \mathcal{H}$, *where a restriction to the symmetric or antisymmetric subspace is tacitly assumed*, and so forth. The Fock space $\mathcal{F}$ is the Hilbert space defined by[2]

$$\mathcal{F} = \mathcal{H}_0 \oplus \mathcal{H}_1 \oplus \mathcal{H}_2 \oplus \mathcal{H}_3 \oplus \cdots . \tag{4.4}$$

The direct sums go on forever! What can that mean?

Let us start by focusing on the harmonic oscillator example. An element of $\mathbf{C}$ is just a complex number, say, ψ_0. An element of $\mathcal{H}_1$ is a square-integrable function $\psi_1(x_1)$. An element of $\mathcal{H}_2$ is a square-integrable function of two variables $\psi(x_1, x_2)$ satisfying either $\psi(x_1, x_2) = \psi(x_2, x_1)$ *or* $\psi(x_1, x_2) = -\psi(x_2, x_1)$, depending upon whether we are dealing with bosons or fermions. An element of $\mathcal{H}_3$ is a square-integrable function of 3 variables, $\psi(x_1, x_2, x_3)$ which is symmetric *or* antisymmetric under any permutation of its arguments. In this way an element $|\Psi\rangle \in \mathcal{F}$ is an infinite sequence:

$$|\Psi\rangle = (\psi_0, \psi_1(x_1), \psi_2(x_1, x_2), \ldots, \psi_n(x_1, x_2, \ldots, x_n), \ldots). \tag{4.5}$$

The scalar product of $|\Psi\rangle$ and $|\Phi\rangle$,

$$|\Phi\rangle = (\phi_0, \phi_1(x_1), \phi_2(x_1, x_2), \ldots, \phi_n(x_1, x_2, \ldots, x_n), \ldots) \tag{4.6}$$

[2]Again, we should form the completion of this vector space using the norm induced by the scalar product.

is an infinite series:

$$\langle\Psi|\Phi\rangle = \psi_0^*\phi_0 + \int_{-\infty}^{\infty} dx_1\, \psi_1^*(x_1)\phi_1(x_1) + \int_{-\infty}^{\infty} dx_1 \int_{-\infty}^{\infty} dx_2\, \psi_2^*(x_1,x_2)\phi_2(x_1,x_2) + \cdots \tag{4.7}$$

We require that this series converges. This is guaranteed if

$$\langle\Psi|\Psi\rangle = |\psi_0|^2 + \int_{-\infty}^{\infty} dx_1\, |\psi_1(x_1)|^2 + \int_{-\infty}^{\infty} dx_1 \int_{-\infty}^{\infty} dx_2\, |\psi_2(x_1,x_2)|^2 + \cdots < \infty \tag{4.8}$$

Evidently, the Hilbert space norm of the functions $\psi_n(x_1, x_2, \ldots, x_n)$ must vanish as $n \to \infty$. As usual, vectors in $\mathcal{F}$ representing states of a quantum system will be normalized, $\langle\Psi|\Psi\rangle = 1$. Notice that this means the individual 1-particle, 2-particle, 3-particle, *etc.,* wave functions will not be normalized to unity.

At this point it is worth previewing the physical interpretation of the Fock state just described. (I will explain how this interpretation arises in a little while.) In the state represented by $|\Psi\rangle$, $|\psi_0|^2$ is the probability that there are no particles present; $|\psi(x_1)|^2$ is the probability density that one particle is present at a displacement x_1; $|\psi(x_1,x_2)|^2$ is the probability density that two particles are present at the displacements x_1 and x_2; $\psi_n(x_1,x_2,\ldots,x_n)|^2$ is the probability density for finding n particles at displacements $x_1, x_2, \ldots, x_n$. The Hilbert space norms of each of the wave functions appearing in (4.8) are then the probabilities for finding $0, 1, 2, \ldots, n, \ldots$ particles, respectively in the state $|\Psi\rangle$. Now you can see how Fock space is the setting for considering systems where the number of particles can vary. It is interesting to note that, statistically speaking, the states $|\Psi\rangle$ need not have a definite number of particles!

4.2 Occupation numbers. Creation and annihilation operators.

I now want to introduce some fundamental tools for working with the Fock space. We begin by focusing on the example of identical harmonic oscillators. Let us use the basis of energy eigenvectors, $|n\rangle \equiv |E_n\rangle$, $n = 0, 1, 2, \ldots$, for the Hilbert space of one oscillator to induce a basis on $\mathcal{F}$. An obvious basis we can use consists of vectors $|\beta\rangle$ of the form

$$|\beta\rangle = \Big(z, |a\rangle, |b\rangle \otimes |c\rangle, |d\rangle \otimes |e\rangle \otimes |f\rangle, \ldots\Big), \quad z \in \mathbf{C}, \quad a, b, c, d, e, f, \cdots = 0, 1, 2, 3, \ldots, \tag{4.9}$$

where I will leave implicit the symmetrization or anti-symmetrization of each element of the sequence. This basis is just fine, but there is a more useful choice we can make. For each choice of $a, b, c, \ldots$ the basis vector (4.9) can itself be viewed as a linear combination of a zero particle state, a 1-particle state, a 2-particle state, and so forth:

$$|\beta\rangle = \Big(1, 0, 0, \ldots\Big) + \Big(0, |a\rangle, 0, 0, \ldots\Big) + \Big(0, 0, |b\rangle \otimes |c\rangle, \ldots\Big) + \ldots \tag{4.10}$$

Again, symmetrization or anti-symmetrization is implicit. Evidently, there is a basis for $\mathcal{F}$ consisting of the following: a state with 0 particles – the *vacuum state*; all states with 1 particle in the various energy eigenstates; all states with two particles in the various energy eigenstates, and so forth. Each element of this basis has a fixed number of particles, say $p = 0, 1, 2, \ldots$. We can uniquely

specify a basis element with p particles by simply specifying how many of the p particles are in the ground state, how many are in the first excited state, and so on. These are the *occupation numbers*. For bosons the occupation numbers can be any non-negative integers. For fermions the occupation numbers can only be 0 or 1. The resulting basis for $\mathcal{F}$ is called the *occupation number* basis. An occupation number basis vector describing p particles, with n_0 in the ground state, with n_1 in the first excited state, and so forth is denoted by

$$|n_0, n_1, n_2, \ldots\rangle, \quad n_0 + n_1 + n_2 + \cdots = p. \tag{4.11}$$

The vacuum state, with no particles present, is usually denoted by $|0\rangle$. Given the energy eigenvector basis (or any other basis we wish to use) for the one particle Hilbert space, the occupation number basis vectors are labeled by an infinite sequence of occupation numbers. The occupation basis is orthonormal. While we used the 1-particle energy eigenvectors to construct the occupation number basis, any basis for the 1-particle Hilbert space can be used.

Two linear operators on $\mathcal{F}$ which are of fundamental utility are the *creation* and *annihilation* operators, denoted by $\hat{a}_k^\dagger$ and $\hat{a}_k$, respectively, where $k = 0, 1, 2, \ldots$. These operators add or remove a particle in the energy eigenstate labeled by the non-negative integer k. To define any linear operator mathematically it suffices to give its action on a basis; here are the definitions of $\hat{a}_k^\dagger$ and $\hat{a}_k$ via the occupation number basis.

Bosons

$$\hat{a}_k|n_0, n_1, \ldots n_k, \ldots\rangle = \sqrt{n_k}|n_0, n_1, \ldots n_k - 1, \ldots\rangle, \tag{4.12}$$

$$\hat{a}_k^\dagger = \sqrt{n_k + 1}|n_0, n_1, \ldots n_k + 1, \ldots\rangle. \tag{4.13}$$

Fermions

$$\hat{a}_k|n_0, n_1, \ldots n_k = 0, \ldots\rangle = 0, \quad a_k|n_0, n_1, \ldots n_k = 1, \ldots\rangle = (-1)^{\alpha(k)}|n_0, n_1, \ldots n_k = 0, \ldots\rangle \tag{4.14}$$

$$\hat{a}_k^\dagger|n_0, n_1, \ldots n_k = 0, \ldots\rangle = (-1)^{\alpha(k)}|n_0, n_1, \ldots n_k = 1, \ldots\rangle, \quad a_k^\dagger|n_0, n_1, \ldots n_k = 1, \ldots\rangle = 0, \tag{4.15}$$

where $(-1)^\alpha$ is ± 1 according to whether the number of occupied 1-particle states lower than k is even/odd. You can check that this rule is satisfied for identical harmonic oscillators if we set

$$\alpha(k) = n_0 + n_1 + \cdots + n_{k-1}. \tag{4.16}$$

In both the boson and fermion cases the result of applying an operator is the zero vector (not the vacuum vector!) if a negative occupation number were to occur. Let us spend a little time understanding the ramifications of all these definitions.

It is straightforward to apply the definitions given above to verify the following fundamental results.

- The creation and annihilation operators are related by Hermitian conjugation:

$$\hat{a}_k^\dagger = (\hat{a}_k)^\dagger. \tag{4.17}$$

- The bosonic operators satisfy the following commutation relations:

$$[\hat{a}_k, \hat{a}_l] = [\hat{a}_k^\dagger, \hat{a}_l^\dagger] = 0, \quad [\hat{a}_k, \hat{a}_l^\dagger] = \delta_{kl}\hat{1}, \tag{4.18}$$

  where the *commutator* is defined as $[\hat{A}, \hat{B}] \equiv \hat{A}\hat{B} - \hat{B}\hat{A}$.

- The fermionic operators satisfy the following anti-commutation relations:

$$\{\hat{a}_k, \hat{a}_l\} = \{\hat{a}_k^\dagger, \hat{a}_l^\dagger\} = 0, \quad \{\hat{a}_k, \hat{a}_l^\dagger\} = \delta_{kl}\hat{1}, \tag{4.19}$$

  where the *anti-commutator* is defined as $\{\hat{A}, \hat{B}\} \equiv \hat{A}\hat{B} + \hat{B}\hat{A}$. Note in particular that $a_k^\dagger a_k^\dagger = 0$, so the occupation number can only be 0 or 1, which is again the Pauli exclusion principle.

- For both bosons and fermions, the *number operator* $\hat{N}_k$ for the 1-particle state labeled by k, defined by

$$\hat{N}_k = \hat{a}_k^\dagger \hat{a}_k, \tag{4.20}$$

  acts on the occupation number basis via

$$\hat{N}_k |n_0, n_1, \ldots n_k, \ldots\rangle = n_k |n_0, n_1, \ldots n_k, \ldots\rangle. \tag{4.21}$$

Evidently, the operator $\hat{N}$ representing the total number of particles is

$$\hat{N} = \sum_k N_k. \tag{4.22}$$

The occupation number basis vectors are eigenvectors of $\hat{N}_k$ and $\hat{N}$. Each of the n-fold tensor product vectors appearing in the superposition (4.10) is an eigenvector of $\hat{N}$ with eigenvalue n. Therefore, as advertised earlier, the wave functions of the form $\psi(x_1, \ldots, x_n)$ appearing in (4.5) are position probability amplitudes for states with n particles.

As mentioned earlier, the state with no particles is the vacuum state, usually denoted $|0\rangle$ (not the zero vector!). It satisfies

$$a_k|0\rangle = 0, \quad \forall\ k, \tag{4.23}$$

and this can be taken to be its definition. A state with a single particle in the one particle state $|k\rangle$ is

$$|0, \ldots, n_k = 1, 0, \ldots\rangle = a_k^\dagger |0\rangle \tag{4.24}$$

If we add a particle to, say, energy state $m < k$, we get the state

$$|0, \ldots, n_m = 1, \ldots, n_k = 1, 0, \ldots\rangle = a_m^\dagger a_k^\dagger |0\rangle \tag{4.25}$$

if we add the particles in the reverse order we get the same vector for bosons and (-1) times the vector in the case of fermions. This implements the convention that the 1-particle states in a multi-particle fermion state are the anti-symmetrization of a state which is in increasing order with respect to energy. A basis for the N-particle subspace of $\mathcal{F}$ is obtained from the vectors

$$a_{k_1}^\dagger a_{k_2}^\dagger \cdots a_{k_N}^\dagger |0\rangle, \quad \forall\ k_1, k_2, \ldots, k_N. \tag{4.26}$$

The union of these bases for each N defines a basis for $\mathcal{F}$. A state vector is any normalized linear combination of these occupation number basis vectors. As such, I remind you that states need

not be eigenvectors of $\hat{N}_k$ or even $\hat{N}$. The number of particles in a state is subject to statistical fluctuations!

Although I have not done it here, one can develop the preceding results in a way which shows that Fock space can be *defined* as a vector space representing the algebra of creation and annihilation operators (4.17)–(4.19). Roughly speaking, once the commutation (or anti-commutation relations) are specified, the Fock space is determined. Here I have simply built the Fock space, postulated the form of the creation and annihilation operators, and deduced some of their properties. For a systematic derivation of creation and annihilation operator formulas see, for example, [2].

4.3 Observables. Field operators.

Here we will explore the construction of operators on $\mathcal{F}$ representing observables of a system with a variable number of particles.

4.3.1 1-particle observables

The simplest observables are the "additive 1-particle" observables, which arise by taking a property which can be assigned to each individual particle and summing it over all particles in the system. We have already seen an example of such an operator: the number operator $\hat{N}$. In our example of a collection of non-interacting oscillators we can measure an energy for each oscillator; the total energy is obtained by summing the result. The operator which represents this observable is

$$\hat{H} = \sum_k E_k \hat{N}_k, \tag{4.27}$$

where $E_k = (k + \frac{1}{2})\hbar\omega$. The occupation number basis we have been using is also the basis of energy eigenvectors:

$$\hat{H}|n_0, n_1, n_2, \ldots\rangle = \left(\sum_k n_k E_k\right) |n_0, n_1, n_2, \ldots\rangle. \tag{4.28}$$

This is possible since $[\hat{H}, \hat{N}_k] = 0$.

The total energy was easy to define since we chose to work in the occupation number basis associated to 1-particle energy eigenstates. To construct operators for other additive 1-particle observables we use a similar strategy: use the basis of eigenvectors of the chosen 1-particle operator to build the Fock basis. For example, let's build the total momentum of the system. We can use the basis of 1-particle momentum eigenvectors $|p\rangle$ to build an occupation number basis for the Fock space. This entails using integrals instead of sums over the basis, and we must be careful to keep in mind that this basis is of the "generalized" type. Recall that, for one particle, the bases of energy eigenvectors $|E_k\rangle$ and momentum eigenvectors $|p\rangle$ are related by

$$|E_k\rangle = \int_{-\infty}^{\infty} dp\, \tilde{u}_k(p)|p\rangle, \quad |p\rangle = \sum_k \tilde{u}_k^*(p)|E_k\rangle, \tag{4.29}$$

where $\tilde{u}_k(p)$ is the momentum space energy eigenfunction

$$\tilde{u}_k(p) = \langle p|E_k\rangle. \tag{4.30}$$

Using these relations, let us define the creation and annihilation operators for particles with momentum p to be $\hat{a}^\dagger(p)$ and $\hat{a}(p)$:

$$\hat{a}^\dagger(p) = \sum_l \tilde{u}_l^*(p)\hat{a}_l^\dagger \tag{4.31}$$

and

$$\hat{a}(p) = \sum_l \tilde{u}_l(p)\hat{a}_l \tag{4.32}$$

The bosonic creation and annihilation operators satisfy the following commutation relations:

$$[\hat{a}(p), \hat{a}(p')] = [\hat{a}^\dagger(p), \hat{a}^\dagger(p')] = 0, \quad [\hat{a}(p), \hat{a}^\dagger(p)] = \delta(p, p')\hat{1}. \tag{4.33}$$

The fermionic operators satisfy the following anti-commutation relations:

$$\{\hat{a}(p), \hat{a}(p')\} = \{\hat{a}^\dagger(p), \hat{a}^\dagger(p')\} = 0, \quad \{\hat{a}(p), \hat{a}^\dagger(p)\} = \delta(p, p')\hat{1}. \tag{4.34}$$

In terms of our original basis (4.9) for $\mathcal{F}$, we have

$$\hat{a}^\dagger(p)(1, 0, 0, \dots) = \sum_k \tilde{u}_k^*(p)a_k^\dagger(1, 0, 0, \dots) = (0, \sum_k \tilde{u}_k^*(p)|E_k\rangle, 0, \dots) = (0, |p\rangle, 0, \dots) \tag{4.35}$$

and

$$\hat{a}^\dagger(p_2)(0, |p_1\rangle, 0, \dots) = \hat{a}^\dagger(p_2)(0, \sum_k \tilde{u}_k^*(p_1)|E_k\rangle, 0, \dots) \tag{4.36}$$

$$= (0, 0, \sum_{k,l} \tilde{u}_k^*(p_1)u_l^*(p_2)(0, 0, \frac{1}{\sqrt{2}}\left[|E_k\rangle \otimes |E_l\rangle \pm |E_l\rangle \otimes |E_k\rangle\right], 0, \dots) \tag{4.37}$$

$$= (0,\, 0,\, \frac{1}{\sqrt{2}}(|p_1\rangle \otimes |p_2\rangle \pm |p_2\rangle \otimes |p_1\rangle), 0, \dots). \tag{4.38}$$

In this way we can build a basis for the Fock space from the 1-particle basis of momentum eigenvectors. Using the same logic as with the energy, the total momentum operator is defined to be

$$\hat{P} = \int_{-\infty}^{\infty} dp\, p\, a^\dagger(p)a(p). \tag{4.39}$$

Using (4.31), (4.32), the total momentum operator can be written in terms of the energy basis as

$$\hat{P} = \int_{-\infty}^{\infty} dp\, p \sum_{k,l} \tilde{u}_k^*(p)u_l(p)\hat{a}_k^\dagger\hat{a}_l \tag{4.40}$$

The expression of $\hat{P}$ in the basis for $\mathcal{F}$ induced by 1-particle energy eigenvectors is a little complicated. A more enlightening thing to do is to repeat the previous type of calculation and express the total energy and momentum operators in terms of the basis of 1-particle position eigenvectors $|x\rangle$ and the corresponding creation and annihilation operators $\hat{a}^\dagger(x)$ and $\hat{a}(x)$, which create and remove a particle at the point x. These operators are known as *field operators*. They can be defined from the creation and annihilation operators for momentum states as follows. From elementary quantum mechanics the relation between the position and momentum eigenvectors is

$$|p\rangle = \int_{-\infty}^{\infty} dx\, \frac{1}{\sqrt{2\pi\hbar}} e^{\frac{i}{\hbar}px}|x\rangle, \quad |x\rangle = \int_{-\infty}^{\infty} dp\, \frac{1}{\sqrt{2\pi\hbar}} e^{-\frac{i}{\hbar}px}|p\rangle \tag{4.41}$$

Using the same logic as with the creation/annihilation operators for particles in momentum states, we define

$$\hat{a}(x) = \int_{-\infty}^{\infty} dp\, \frac{1}{\sqrt{2\pi\hbar}} e^{\frac{i}{\hbar}px} \hat{a}(p), \quad \hat{a}^\dagger(x) = \int_{-\infty}^{\infty} dx\, \frac{1}{\sqrt{2\pi\hbar}} e^{-\frac{i}{\hbar}px} \hat{a}^\dagger(p), \tag{4.42}$$

which implies

$$\hat{a}(p) = \int_{-\infty}^{\infty} dx\, \frac{1}{\sqrt{2\pi\hbar}} e^{-\frac{i}{\hbar}px} \hat{a}(x), \quad \hat{a}^\dagger(p) = \int_{-\infty}^{\infty} dx\, \frac{1}{\sqrt{2\pi\hbar}} e^{\frac{i}{\hbar}px} \hat{a}^\dagger(x). \tag{4.43}$$

The field operators satisfy the following commutation relations:

$$[\hat{a}(x), \hat{a}(x')] = [\hat{a}^\dagger(x), \hat{a}^\dagger(x')] = 0, \quad [\hat{a}(x), \hat{a}^\dagger(x)] = \delta(x, x')\hat{1}. \tag{4.44}$$

The fermionic operators satisfy the following anti-commutation relations:

$$\{\hat{a}(x), \hat{a}(x')\} = \{\hat{a}^\dagger(x), \hat{a}^\dagger(x')\} = 0, \quad \{\hat{a}(x), \hat{a}^\dagger(x)\} = \delta(x, x')\hat{1}. \tag{4.45}$$

The field operators add particles in (generalized) position eigenstates to the occupation number basis vectors. For example,

$$a^\dagger(x)(1, 0, 0, \dots) = (0, |x\rangle, 0, 0, \dots). \tag{4.46}$$

Let us express the total momentum in terms of the field operators. Substituting (4.43) into the original formula (4.39) gives:

$$\hat{P} = \frac{1}{2\pi\hbar} \int_{-\infty}^{\infty} dp \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dy\, p\, e^{\frac{i}{\hbar}p(y-x)} \hat{a}^\dagger(y)\hat{a}(x) \tag{4.47}$$

$$= \int_{-\infty}^{\infty} dx\, \hat{a}^\dagger(x) \left(\frac{\hbar}{i}\frac{d}{dx}\right) \hat{a}(x). \tag{4.48}$$

Here we have used the Fourier definition of the derivative of the delta function:

$$\frac{1}{2\pi\hbar} \int_{-\infty}^{\infty} dp\, p\, e^{\frac{i}{\hbar}p(y-x)} = -\frac{1}{2\pi i}\frac{\partial}{\partial x} \int_{-\infty}^{\infty} dp\, e^{\frac{i}{\hbar}p(y-x)} = -\frac{\hbar}{i}\frac{\partial}{\partial x}\delta(y-x), \tag{4.49}$$

in conjunction with the delta-function identity[3]

$$\int_{-\infty}^{\infty} dx\, \frac{\partial}{\partial x}\delta(y-x) f(x) = -f'(y). \tag{4.50}$$

Let us next express the total energy operator in terms of the field operators. Start with the relation between energy eigenvectors and position eigenvectors:

$$|E_k\rangle = \int_{-\infty}^{\infty} dx\, u_k(x)|x\rangle, \quad |x\rangle = \sum_k u_k^*(x)|E_k\rangle. \tag{4.51}$$

[3]It is possible to prove that such identities for functions extend to the field operators by a limiting process such as we discussed earlier.

As before, this implies

$$\hat{a}_k = \int_{-\infty}^{\infty} dx\, u_k^*(x)\hat{a}(x), \quad \hat{a}_k^\dagger = \int_{-\infty}^{\infty} dx\, u_k(x)\hat{a}^\dagger(x), \tag{4.52}$$

and

$$\hat{a}(x) = \sum_n \hat{a}_n u_n(x), \quad \hat{a}^\dagger(x) = \sum_n \hat{a}_n^\dagger u_n^*(x). \tag{4.53}$$

We can substitute these relations into (4.27), then use the eigenfunction condition

$$\left(-\frac{\hbar^2}{2m}\frac{d^2}{dx^2} + \frac{1}{2}m\omega^2 x^2\right) u_k(x) = E_k u_k(x), \tag{4.54}$$

and the self-adjointness of the Hamiltonian

$$\int_{-\infty}^{\infty} dx\, f^*(x)\left(-\frac{\hbar^2}{2m}\frac{d^2}{dx^2} + \frac{1}{2}m\omega^2 x^2\right) g(x) = \int_{-\infty}^{\infty} dx\, g(x)\left(-\frac{\hbar^2}{2m}\frac{d^2}{dx^2} + \frac{1}{2}m\omega^2 x^2\right) f^*(x), \tag{4.55}$$

which can be extended to the field operators with the substitution of Hermitian adjoint (†) for complex conjugation, to find

$$\hat{H} = \int_{-\infty}^{\infty} dx\, a^\dagger(x)\left(-\frac{\hbar^2}{2m}\frac{d^2}{dx^2} + \frac{1}{2}m\omega^2 x^2\right) a(x). \tag{4.56}$$

Are you starting to see a pattern? The operator $\hat{A}$ representing the total value of an additive 1-particle observable A_1 can be constructed by (i) expressing $\hat{A}_1 = A_1(\hat{x}, \hat{p})$ for a single oscillator; (ii) formally constructing the expectation value of $\hat{A}_1$ using a wave functions $\psi(x)$ and its conjugate $\psi^*(x)$; (iii) converting the wave function and its conjugate to the field operators $a(x)$ and $a^\dagger(x)$, respectively.

4.3.2 2-particle observables

There are observables which necessarily involve 2 or more particles. The most important of these is the potential energy of interaction between particles. Interactions between pairs of particles can be introduced via potential energies that depend upon the locations of any two particles. Let us denote the potential energy between particles at x and x' by $V(x, x')$. For identical particles we must have $V(x, x') = V(x', x)$. For example, one could couple oscillators with displacements x and x' via a potential of the form

$$V(x, x') = \alpha e^{-\beta|x-x'|}, \quad \alpha = const. \tag{4.57}$$

The operator $\hat{V}$ which represents the total value of this observable can be found by building an operator that determines which points x have a particle, applying the formula above for all pairs of such points, and then adding up the results. For example, suppose we have a state with 3 particles at locations x_1, x_2, x_3. The potential energy observable for this state would be

$$V = V(x_1, x_2) + V(x_1, x_3) + V(x_2, x_3). \tag{4.58}$$

More generally, for N particles at locations x_i, $i = 1, 2, \dots N$, we have

$$V = \sum_{i<j=1}^{N} V(x_i, x_j) = \frac{1}{2} \sum_{i\neq j=1}^{N} V(x_i, x_j). \tag{4.59}$$

The operator

$$\hat{V} = \frac{1}{2} \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dx'\, V(x, x') \hat{a}^\dagger(x') \hat{a}^\dagger(x) \hat{a}(x) \hat{a}(x') \tag{4.60}$$

represents this observable in the sense that a state with particles at $x_1, x_2, \dots, x_N$, represented by the vector

$$|x_1, x_2, \dots, x_N\rangle = a^\dagger(x_1) a^\dagger(x_2) \dots a^\dagger(x_N)|0\rangle, \tag{4.61}$$

yields the potential (4.59) with certainty:

$$\hat{V}|x_1, x_2, \dots, x_N\rangle = \left[\frac{1}{2} \sum_{i\neq j=1}^{N} V(x_i, x_j) \right] |x_1, x_2, \dots, x_N\rangle. \tag{4.62}$$

We saw earlier that the additive one particle observables are represented by operators which are quadratic in field operators. Such observables describe aspects of non-interacting particles. As you can see from (4.60), 2-particle interactions involve a potential energy quartic in field operators. 3-particle interactions would involve polynomials of degree 6 in the field operators.

4.3.3 Field operators and wave functions

The field operators can be used to recover the more traditional description of a many-body system (with a fixed number of particles) in terms of wave functions. To begin, let $|\Psi_1\rangle$ be a Fock state containing 1-particle. Viewing state vectors in $\mathcal{F}$ as a sequence of vectors according to (4.4), we can write

$$|\Psi_1\rangle = (0, |\psi_1\rangle, 0, 0, \dots), \tag{4.63}$$

where $|\psi_1\rangle \in \mathcal{H}_1$. Using

$$|\psi_1\rangle = \int_{-\infty}^{\infty} dy\, \psi_1(y)|y\rangle, \tag{4.64}$$

we can write

$$|\Psi_1\rangle = \int_{-\infty}^{\infty} dy\, \psi_1(y) \hat{a}^\dagger(y)|0\rangle. \tag{4.65}$$

Evidently, the wave function $\psi_1(x)$ characterizing $|\Psi_1\rangle$ can be obtained from the formula

$$\psi_1(x) = \langle 0|\hat{a}(x)|\Psi_1\rangle, \tag{4.66}$$

as you can verify using the (anti-)commutation relations of the field operators and $\hat{a}(x)|0\rangle = 0$. Next consider a 2-particle state

$$|\Psi_2\rangle = (0, 0, |\psi_2\rangle, 0, \dots), \quad |\psi_2\rangle \in \mathcal{H}_1 \otimes \mathcal{H}_1, \tag{4.67}$$

where we take the symmetric or anti-symmetric part of the tensor product, as usual. As you know, we can express $|\psi_2\rangle$ in terms of a wave function of 2 variables,

$$|\psi_2\rangle = \int_{-\infty}^{\infty} dx_2 \int_{-\infty}^{\infty} dx_1 \psi(x_1, x_2) \frac{1}{\sqrt{2}} (|x_1\rangle \otimes |x_2\rangle \pm |x_2\rangle \otimes |x_1\rangle), \tag{4.68}$$

where $\psi(x_1, x_2)$ is symmetric or anti-symmetric under particle interchange. In this 2-particle case we have

$$|\Psi_2\rangle = \frac{1}{\sqrt{2}} \int_{-\infty}^{\infty} dx_2 \int_{-\infty}^{\infty} dx_1 \psi(x_1, x_2) a^\dagger(x_2) a^\dagger(x_1) |0\rangle. \tag{4.69}$$

Given $|\Psi_2\rangle$ we can recover the wave function via

$$\psi(x_1, x_2) = \frac{1}{\sqrt{2}} \langle 0| a(x_1) a(x_2) |\Psi_2\rangle. \tag{4.70}$$

Notice that the symmetry properties of the wave function are correctly implemented by the commutation or anti-commutation properties of the field operators. In general, we can represent an n-particle state $|\Psi\rangle$ by a function $\psi(x_1, x_2, \ldots, x_n)$ with appropriate symmetry properties; this wave function is given by

$$\psi(x_1, x_2, \ldots, x_n) = \frac{1}{\sqrt{n!}} \langle 0| a(x_1) a(x_2) \cdots a(x_n) |\Psi\rangle. \tag{4.71}$$

In this way one can view a wave function $\psi(x_1, x_2, \ldots, x_n)$ as the component of the state vector along the basis element corresponding to having a particle at each $x_1, x_2, \ldots, x_n$,

$$\frac{1}{\sqrt{n!}} a^\dagger(x_n) a^\dagger(x_{n-1}) \cdots a^\dagger(x_1) |0\rangle.$$

Of course, a state vector in $\mathcal{F}$ is in general a superposition of states with various numbers of particles. One can extract the wave function for each of the elements of this superposition from (4.71) for different choices of n. In this way the traditional wave function representation of states of a system with a fixed number of (identical) particles is derived from the more general Fock space formalism, and we see another key aspect of the field operator.

We can also see how traditional formulas for expectation values in terms of wave functions for a fixed number of particles arise from expectation values in Fock space.[4] For example, let us compute the expectation value of momentum for a single oscillator in the state described by the wave function $\psi(x) \in \mathcal{H}_1$. The Fock state $|\Psi\rangle$ representing this situation is

$$|\Psi\rangle = \int_{-\infty}^{\infty} dx\, \psi(x) \hat{a}^\dagger(x) |0\rangle. \tag{4.72}$$

[4] I remind you that all the physical output of quantum mechanics can be obtained via expectation values.

Using (4.48), the expectation value of total momentum is

$$\begin{aligned}\langle P\rangle &= \langle\Psi|\hat{P}|\Psi\rangle \\ &= \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dy \int_{-\infty}^{\infty} dz\, \psi^*(x)\psi(z)\langle 0|\hat{a}(x)\hat{a}^\dagger(y)\left(\frac{\hbar}{i}\frac{d}{dy}\hat{a}(y)\right)a^\dagger(z)|0\rangle \\ &= \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dy \int_{-\infty}^{\infty} dz\, \psi^*(x)\psi(z)\langle 0|\hat{a}(x)\hat{a}^\dagger(y)|0\rangle\frac{\hbar}{i}\frac{d}{dy}\delta(y,z) \\ &= \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dy \int_{-\infty}^{\infty} dz\, \psi^*(x)\psi(z)\delta(x,y)\frac{\hbar}{i}\frac{d}{dy}\delta(y,z) \\ &= \int_{-\infty}^{\infty} dx\, \psi^*(x)\left(\frac{\hbar}{i}\frac{d}{dx}\right)\psi(x). \end{aligned} \tag{4.73}$$

To get from the second line to the third line in (4.73) I used the simple identity

$$\left(\frac{\hbar}{i}\frac{d}{dy}\hat{a}(y)\right)a^\dagger(z) = \frac{\hbar}{i}\frac{d}{dy}\left(\hat{a}(y)a^\dagger(z)\right) = \frac{\hbar}{i}\frac{d}{dy}\left(\pm a^\dagger(z)\hat{a}(y) + \delta(y,z)\hat{1}\right). \tag{4.74}$$

4.4 Time evolution of the field operators

So far we have been investigating the properties of states and observables in Fock space which are independent of time evolution. Now we take account of dynamics. For simplicity we just consider a system of identical, non-interacting oscillators.[5]

The creation and annihilation operators associated to various types of 1-particle bases suffice to build up all observables, so we describe the time evolution of the system by determining how these operators should evolve in time in the Heisenberg picture. We take the Hamiltonian to be the total energy of the system of oscillators expressed in terms of the creation and annihilation operators defined by the 1-particle energy eigenvectors. This is defined in (4.27), repeated here for your convenience:

$$\hat{H} = \sum_{k=0}^{\infty} E_k\, \hat{N}_k. \tag{4.75}$$

You can easily check that

$$[\hat{a}_j, \hat{N}_k] = \hat{a}_k\delta_{jk}\hat{1}. \tag{4.76}$$

The Heisenberg operator $\hat{a}_j(t)$ therefore satisfies

$$i\hbar\frac{d}{dt}\hat{a}_j(t) = E_j\,\hat{a}_j(t), \quad E_j = (j+\frac{1}{2})\hbar\omega, \tag{4.77}$$

with the solution

$$\hat{a}_j(t) = e^{-\frac{i}{\hbar}E_j t}\hat{a}_j. \tag{4.78}$$

Similarly,

$$\hat{a}_j^\dagger(t) = e^{\frac{i}{\hbar}E_j t}\hat{a}_j^\dagger. \tag{4.79}$$

[5]Taking account of interactions and developing strategies for analyzing their effects is in many ways the principal challenge in quantum field theory. Unfortunately, to go into all that is beyond the scope of this introductory course.

Let us next consider the time evolution of the "field operators" $\hat{a}(x)$ and $\hat{a}^\dagger(x)$, introduced earlier. Recall from (4.56) that the Hamiltonian can be expressed in terms of them as

$$\hat{H} = \int_{-\infty}^{\infty} dy\, a^\dagger(y) \left(-\frac{\hbar^2}{2m}\frac{d^2}{dy^2} + \frac{1}{2}m\omega^2 y^2 \right) a(y). \tag{4.80}$$

To compute the Heisenberg equations we need

$$[\hat{a}(x), \hat{H}] = \int_{-\infty}^{\infty} dy\, [\hat{a}(x), \hat{a}^\dagger(y)] \left(-\frac{\hbar^2}{2m}\frac{d^2}{dy^2} + \frac{1}{2}m\omega^2 y^2 \right) \hat{a}(y) \tag{4.81}$$

$$= \int_{-\infty}^{\infty} dy\, \delta(x,y) \left(-\frac{\hbar^2}{2m}\frac{d^2}{dy^2} + \frac{1}{2}m\omega^2 y^2 \right) \hat{a}(y) \tag{4.82}$$

$$= \left(-\frac{\hbar^2}{2m}\frac{d^2}{dx^2} + \frac{1}{2}m\omega^2 x^2 \right) \hat{a}(x). \tag{4.83}$$

The Heisenberg equations for the field operators $\hat{a}(x,t)$ and $\hat{a}^\dagger(x,t)$ at time t are therefore

$$i\hbar\frac{\partial}{\partial t}\hat{a}(x,t) = \left(-\frac{\hbar^2}{2m}\frac{d^2}{dx^2} + \frac{1}{2}m\omega^2 x^2 \right) \hat{a}(x,t), \tag{4.84}$$

$$-i\hbar\frac{\partial}{\partial t}\hat{a}^\dagger(x,t) = \left(-\frac{\hbar^2}{2m}\frac{d^2}{dx^2} + \frac{1}{2}m\omega^2 x^2 \right) \hat{a}^\dagger(x,t), \tag{4.85}$$

These equations are formally identical to the Schrödinger equation for a harmonic oscillator wave function! Of course, the meaning of the Heisenberg equations (4.84), (4.85) is very different from that of the Schrödinger equation. The role of the wave function is now being played by a family of linear operators on Fock space. The operator $\hat{a}^\dagger(x,t)$ adds to the state vector a particle/oscillator at the time t and displacement x while $\hat{a}(x,t)$ removes a particle/oscillator at the time t and displacement x. The solution to (4.84) is

$$\hat{a}(x,t) = \sum_n \hat{a}_n e^{-\frac{i}{\hbar}E_n t} u_n(x) = \sum_n \hat{a}_n(t) u_n(x). \tag{4.86}$$

This recovers the relation (4.53) between creation/annihilation operators for particles in energy eigenstates and the field operators at time t in the Heisenberg picture.

This, then, is where you are inexorably led if you take a given quantum system (*e.g.,* the oscillator) and then consider a composite system consisting of any number of identical copies of that original system: the wave function for the single system gets promoted to a *field operator* which creates and destroys particles. Indeed, one often uses the notation $\hat{a}(x,t) \equiv \hat{\psi}(x,t)$. You may recall that the original construction of the single quantum system (*e.g.,* the oscillator) can be viewed as taking a classical system parametrized by coordinates and momenta and turning these dynamical variables into operators on wave functions. This is sometimes called "first quantization". When one takes the resulting quantum system and then considers its many-body generalization in Fock space the wave function itself gets turned into an operator and one sometimes speaks of "second quantization". The "first quantization" of a particle was created by physicists in the early 20^{th} century to explain experimental results at atomic length scales. It was not derived from some other theory, but it was created to explain observations. "Second quantization", on the other hand, follows from applying the standard "first quantization" rules to a system containing an indefinite number of particles. As such, "second quantization" *is* derived from fundamental principles. As the mathematical physicist Edward Nelson famously said: "First quantization is a mystery, but second quantization is a functor."

4.5 General formalism

Let me summarize our various investigations and put them into a more general context. Consider a quantum particle described by a Hilbert space $\mathcal{H}$. There will exist complete sets of commuting observables; fix one set $\hat{A}_1, \hat{B}_1, \ldots$ with its orthonormal basis of eigenvectors $|e_j\rangle$, $j = 1, 2, \ldots$. While I will use a notation adapted to discrete spectra one can make the usual transition to continuous variables to handle continuous spectra and generalized eigenvectors. The goal now is to construct the quantum theory of any system consisting of these particles. I will focus attention on describing a system of non-interacting particles and I will just mention how to incorporate interactions at the end.

Define the Fock space $\mathcal{F}$ from the direct sum of the symmetric or antisymmetric tensor products of $\mathcal{H}$ according to (4.4) depending upon whether the particles are bosons or fermions, respectively. Define the occupation number basis for particles in the 1-particle states $|e_j\rangle$. Define creation and annihilation operators $\hat{a}_i^\dagger$, $\hat{a}_i$ for a particle in the state $|e_i\rangle$ according to (4.12) – (4.15). Operators representing observables are typically constructed as polynomials in the creation and annihilation operators. For example, the operator representing the number of particles in the state $|e_k\rangle$ is given by

$$\hat{N}_k = \hat{a}_k^\dagger \hat{a}_k, \tag{4.87}$$

and the operator representing the total number of particles is

$$\hat{N} = \sum_k \hat{a}_k^\dagger \hat{a}_k. \tag{4.88}$$

If $\hat{A}_1$ is one of the operators on $\mathcal{H}$ for which $|e_j\rangle$ is a basis of eigenvectors,

$$\hat{A}_1 |e_j\rangle = a_j |e_j\rangle, \tag{4.89}$$

the total value for this observable (at least for non-interacting particles) is represented by the operator

$$\hat{A} = \sum_j a_j \hat{N}_j. \tag{4.90}$$

If $|f_i\rangle$, is another orthonormal basis for $\mathcal{H}$, one can of course repeat the constructions summarized above. In particular, one can construct the associated creation and annihilation operators $\hat{b}_i$, $\hat{b}_i^\dagger$. To relate the two resulting frameworks one uses the relation between the two bases,

$$|e_i\rangle = \sum_j u_{ij} |f_j\rangle, \quad |f_i\rangle = \sum_j u_{ji}^* |e_j\rangle, \tag{4.91}$$

where u_{ij} are the components of a unitary matrix,

$$u_{ij} = \langle f_j | e_i \rangle, \quad \sum_i u_{ij} u_{ik}^* = \delta_{jk}. \tag{4.92}$$

The creation and annihilation operators in the two frameworks are related by

$$b_k = \sum_l u_{kl} a_l, \quad b_k^\dagger = \sum_l u_{kl}^* a_l^\dagger. \tag{4.93}$$

$$a_k = \sum_j u^*_{jk} b_j, \quad a^\dagger_k = \sum_j u_{jk} b^\dagger_j. \tag{4.94}$$

From these relations one can transform back and forth between constructions built upon either basis for $\mathcal{H}$. Different bases for $\mathcal{H}$ provide states which endow the particles with different observable properties, *e.g.,* one basis might yield a Fock basis with particles having statistically definite energies, while another basis might endow the particles with definite momenta. However, one important feature is preserved under a change of basis: the total number of particles. This follows from the identity

$$\sum_k \hat{b}^\dagger_k \hat{b}_k = \sum_{k,l,m} u^*_{kl} u_{km} \hat{a}^\dagger_l \hat{a}_m = \sum_{l,m} \delta_{ml} a^\dagger_l \hat{a}_m = \sum_m a^\dagger_m \hat{a}_m \tag{4.95}$$

In particular, suppose $|N\rangle$ is an eigenvector of the total number operators $\sum_k \hat{a}^\dagger_k \hat{a}_k$ with eigenvalue N, so that this state has N particles in the 1-particle states $|e_i\rangle$, $i = 1, 2, \ldots$, with certainty. Then from (4.95) it follows that there are N particles in the 1-particles states $|f_k\rangle$, $k = 1, 2, \ldots$, with certainty.

As a particularly important example of this formalism for changing 1-particle bases, let $|x\rangle$ represent position eigenvectors for a spinless particle. Here x can represent position in $\mathbf{R}^n$ for any value of n, although $n = 1, 2, 3$ would be most common.[6] The wave functions in position space representing the eigenvectors $|e_i\rangle$ are

$$u_i(x) = \langle x | e_i \rangle. \tag{4.96}$$

Define field operators $\hat{a}(x)$, $\hat{a}^\dagger(x)$ according to:

$$\hat{a}(x) = \sum_k u_k(x) \hat{a}_k, \quad \hat{a}^\dagger(x) = \sum_k u^*_k(x) \hat{a}^\dagger_k \tag{4.97}$$

$$\hat{a}_k = \int_{\text{all space}} dx\, u^*_k(x) \hat{a}(x), \quad \hat{a}^\dagger_k = \int_{\text{all space}} dx\, u_k(x) \hat{a}^\dagger(x). \tag{4.98}$$

These operators obey the (anti)commutation relations (4.44) and (4.45). They act by adding and subtracting a particle at the location x.

The field operators are related to position wave functions via the generalization of (4.71). Let $|\Psi\rangle \in \mathcal{F}$ be an n-particle state. The corresponding n-particle wave function is

$$\psi(x_1, \cdots, x_n) = \langle 0 | \frac{1}{\sqrt{n!}} a(x_1) \cdots a(x_n) | \Psi \rangle. \tag{4.99}$$

Let an additive 1-particle observable A_1 be represented by the operator $\hat{A}_1 \colon \mathcal{H} \to \mathcal{H}$. The extension of A_1 to a system consisting of these particles is represented in Fock space by the operator

$$\hat{A} = \int_{\text{all space}} dx\, a^\dagger(x) \hat{A}_1 \hat{a}(x), \tag{4.100}$$

where it is understood that the representation of $\hat{A}_1$ as a differential operator on functions of x is being used to evaluate $\hat{A}_1 \hat{a}(x)$. The observable A represents the total value of A_1.

[6] Up to this point in this section, we really could be considering multiple copies of any quantum system described by $\mathcal{H}$. By invoking a basis of position eigenvectors we commit to a quantum system representing a particle.

We can recover the usual quantum mechanical formula for the expectation value of the 1-particle observable A_1 by calculating the expectation value of the observable A for the total value in a state consisting of a single particle. Let $|\Psi\rangle \in \mathcal{F}$ be a 1-particle Fock state. We then have

$$\langle A\rangle = \langle\Psi|\hat{A}|\Psi\rangle = \int_{\text{all space}} dx\, \langle\Psi|a^\dagger(x)\hat{A}_1\hat{a}(x)|\Psi\rangle = \int_{\text{all space}} dx\, \psi^*(x)\hat{A}_1\psi(x), \tag{4.101}$$

where the correspondence (4.99) was used. This same sort of result can be extended to a state consisting of any given number of particles. Since all physical predictions can be expressed as expectation values, this shows that the second quantized formalism includes the first quantized formalism and reduces to it when one restricts attention to states with a fixed number of particles.

In the case where the particles are non-interacting, the Hamiltonian operator $\hat{H}$ is just the sum of the 1-particle energies, which are represented by the 1-particle Hamiltonian $\hat{H}_1$:

$$\hat{H} = \int_{\text{all space}} dx\, a^\dagger(x)\hat{H}_1\hat{a}(x). \tag{4.102}$$

The Hamiltonian defines the time dependence of operators representing observables in the Heisenberg picture. Since everything can be constructed using the field operators, let us focus is on their time dependence. The equations of motion – the *field equations* – are

$$i\hbar\frac{\partial}{\partial t}\hat{a}(x,t) = [\hat{a}(x),\hat{H}](t) = \hat{H}_1\hat{a}(x,t). \tag{4.103}$$

Again, this is formally the Schrödinger equation with the wave function having been promoted to linear operator on Fock space. The field equation is a linear equation for the field operators at time t. As such, it is easy to handle mathematically and one has the same control over the theory of non-interacting particles as represented by quantum fields as one would have with a single particle described by the Hamiltonian $\hat{H}_1$. In particular, if $\hat{H}_1$ does not depend upon time and $u_k(x)$, $k = 1, 2, \ldots$, are a basis of energy eigenfunctions for $\hat{H}_1$,

$$\hat{H}_1 u_k(x) = E_k u_k(x), \tag{4.104}$$

then

$$\hat{a}(x,t) = \sum_n e^{-\frac{i}{\hbar}E_n t}u_k(x)\hat{a}_n, \tag{4.105}$$

where $\hat{a}_n$ are (Schrödinger picture) annihilation operators associated to the 1-particle energy basis.

If one were to include interactions between the particles then the Hamiltonian would include additional terms such as the 2-body interaction $\hat{V}$ described in the last section. Such terms will be fourth (or higher) degree polynomials in the field operators. You can easily check that such terms in the Hamiltonian will add *non-linear* terms to the field equations. These non-linearities make the interacting theory more challenging mathematically and quite rich physically. So far, all interactions in nature can be successfully described by such non-linearities.

4.6 Relation to the Hilbert space of quantum normal modes

Let me continue to use the general description of Fock space from the previous section to make one of the principal points of this chapter – indeed, this will be one of the Big Messages of the whole text.

You may have already noticed that, for a given 1-particle state $|e_k\rangle$, the creation and annihilation operators $\hat{a}_k^\dagger$, $\hat{a}_k$ and the number operator $\hat{N}_k$ are *mathematically* identical to that of a single fictitious harmonic oscillator,[7] assuming the particles are bosons. In particular, the creation and annihilation operators for a particle in state $|e_k\rangle$ satisfy the commutation relations of the form (2.45) and (2.50), as you can see from (4.18) and (4.76). In this identification, adding/removing a particle in the state $|e_k\rangle$ is the same as adding/removing a quantum of energy to the fictitious oscillator. The natural angular frequency of the fictitious oscillator is unspecified by this correspondence and can be anything you like. (This frequency would be used to define the Hamiltonian for the "oscillator" and recover an interpretation in terms of displacements.) It follows that the subspace $\mathcal{F}_k \subset \mathcal{F}$ spanned by all occupation number states associated to a fixed 1-particle state $|e_k\rangle$, with occupation numbers for all other 1-particle states set to zero, is isomorphic[8], to the Hilbert space of a single oscillator, call it $\mathcal{H}_k$. For a given choice of mass and angular frequency, the correspondence between the states can be made via

$$|E_m\rangle \quad \longleftrightarrow \quad |0, 0, \ldots, n_k = m, 0, \ldots\rangle, \quad m = 0, 1, 2, \ldots, \tag{4.106}$$

where $|E_m\rangle$ are a basis of energy eigenvectors for the fictitious oscillator. This mapping between orthonormal bases defines the isomorphism. The isomorphism can be extended to identify the bosonic Fock space with the tensor product of single oscillator Hilbert spaces,

$$\mathcal{F} \approx \mathcal{H}_1 \otimes \mathcal{H}_2 \otimes \cdots \otimes \mathcal{H}_m \otimes \cdots \tag{4.107}$$

by the correspondence

$$|E_{a_1}\rangle \otimes |E_{a_2}\rangle \otimes \cdots \otimes |E_{a_m}\rangle \otimes \cdots \quad \longleftrightarrow \quad |n_1 = a_1, n_2 = a_2, \ldots, n_m = a_m, \ldots\rangle. \tag{4.108}$$

Of course, $\mathcal{H}_1 \otimes \mathcal{H}_2 \otimes \cdots$ is just the Hilbert space of an arbitrarily large number of uncoupled oscillators. In this way the Fock space of bosons – for any kind of bosonic system – can always be put into mathematical correspondence with that of an infinite collection of uncoupled harmonic oscillators (with any frequencies you like). Any system described by normal modes of vibration, in the limit where the number of degrees of freedom becomes arbitrarily large, will have a description in terms of a Fock space of bosons. This mathematical feature of the Fock space will be of fundamental interest to us later when we obtain the theory of photons from the quantum electromagnetic field. More generally, this correspondence is ultimately the reason why one can identify the physics of "free quantum fields" with the physics of non-interacting identical particles.

4.7 Problems

1. Show that the occupation number basis is orthonormal.
2. Using the definitions (4.12)–(4.15), verify the properties (4.17)–(4.21).
3. Using (4.31) and (4.32) verify (4.33) and (4.34).

[7] Do not confuse this mathematical "fictitious" oscillator with the oscillators used to illustrate the formalism in the previous section. Here the particles being considered could be any kind of bosonic identical particles and the oscillator interpretation is a new, interesting mathematical feature of Fock space.

[8] The term "isomorphic" means there is a one to one, onto mapping between the two vector spaces such that the scalar product between the vectors is preserved. With a suitable labeling of vectors the two spaces can be identified.

4. Compute the total energy operator (4.27) in terms of the creation and annihilation operators of particles with a given momentum, $\hat{a}(p)$ and $\hat{a}^\dagger(p)$.

5. Verify (4.48) and (4.56).

6. Show that the total momentum and energy given in (4.48) and (4.56) satisfy

$$[\hat{P}, \hat{H}] = -i\hbar m\omega^2 \int_{-\infty}^{\infty} dx\, x\, \hat{a}^\dagger(x)\hat{a}(x).$$

Remark: This is a special case of the following identity, which you may want to prove instead.

If $\hat{A}_1$ and $\hat{B}_2$ are operators on the 1-particle Hilbert space with commutator $\hat{C}_1 \equiv [\hat{A}_1, \hat{B}_1]$, and with Fock space generalizations

$$\hat{A} = \int_{-\infty}^{\infty} dx\, \hat{a}^\dagger(x)(\hat{A}_1 a)(x), \quad \hat{B} = \int_{-\infty}^{\infty} dx\, \hat{a}^\dagger(x)(\hat{B}_1 a)(x), \quad \hat{C} = \int_{-\infty}^{\infty} dx\, \hat{a}^\dagger(x)(\hat{C}_1 a)(x),$$

then

$$[\hat{A}, \hat{B}] = \hat{C}.$$

7. Define number operators $\hat{a}^\dagger(x)\hat{a}(x)$ and $\hat{a}^\dagger(p)\hat{a}(p)$ which count particles at a given displacement x and given momentum p, respectively. Show that the integral over all x or p of the respective operators yields the operator representing the total number of particles in all energy states.

8. Verify (4.62) in the case of three particles.

9. For a system of identical particles in one dimension, define position and momentum operators

$$\hat{X} = \int_{-\infty}^{\infty} dx\, \hat{a}^\dagger(x) x \hat{a}(x), \quad \hat{P} = \int_{-\infty}^{\infty} dp\, \hat{a}^\dagger(p) p \hat{a}(p).$$

Show that

$$[\hat{X}, \hat{P}] = i\hbar \hat{N},$$

where $\hat{N}$ is the operator representing the total number of particles. What is the physical interpretation of the observables X and P corresponding to these operators? (*Hint:* Have a look at the remark in problem 6.)

10. Let $\hat{A}_1$ and $\hat{B}_1$ be operators on a Hilbert space $\mathcal{H}$ representing two observables for a particle. Let their eigenvectors be $|a_i\rangle$ and $|b_i\rangle$, respectively. Define occupation number bases corresponding to each 1-particle basis exactly as we did with the energy eigenvectors. Define the creation and annihilation operators for particles in the states $|a_i\rangle$ and $|b_i\rangle$ as in (4.12)–(4.15). Finally, define the operators representing the total observables A and B for a system of particles and express them in terms of the two types of creation and annihilation operators.

11. Verify (4.70).

12. The energy of a free, relativistic particle of rest mass m and momentum $\vec{p}$ (in a given inertial reference frame) corresponds to the Hamiltonian

$$H = \sqrt{p^2c^2 + m^2c^4},$$

where c is the speed of light. Write down the Schrödinger equation defined by this Hamiltonian in the momentum representation and find its general solution. Take the Fourier transform to get the solution in the position representation. Show that the general solution $\psi(t, \vec{x})$ also satisfies the *Klein-Gordon equation*:

$$-\frac{1}{c^2}\frac{\partial \psi}{\partial t^2} + \nabla^2\psi - \frac{m^2c^2}{\hbar^2}\psi = 0.$$

What are the energy eigenfunctions in the momentum representation? Give a formula for the total energy of a system of such particles at time t in terms of creation and annihilation operators for particles with a given momentum at time t.

13. Verify (4.101). Compute the expectation value of $\hat{A}$ in a 2-particle state and show that you get the desired result in terms of the wave functions and operator $\hat{A}_1$ in the first quantized formalism.

14. Consider a fermion which can be in two states $|\pm\rangle$, these states forming an orthonormal basis for its Hilbert space $\mathcal{H}$. A system consisting of such particles is described by a Hamiltonian operator on the fermionic Fock space given by

$$\hat{H} = \epsilon(\hat{a}_+^\dagger \hat{a}_+ + \hat{a}_-^\dagger \hat{a}_-) + \lambda(\hat{a}_+^\dagger \hat{a}_-^\dagger + \hat{a}_- \hat{a}_+).$$

Show that the number of particles is not preserved in time. Indeed, the Hamiltonian includes terms which create/destroy pairs of particles. Define new operators

$$\hat{b}_+ = u\hat{a}_+ - v\hat{a}_-^\dagger, \quad \hat{b}_- = u\hat{a}_- + v\hat{a}_+^\dagger, \quad u, v \in \mathbf{R}, \quad u^2 + v^2 = 1$$

and show that they satisfy the anti-commutation relations of creation and annihilation operators. Show that there exist u and v such that the Hamiltonian can be written in the form

$$\hat{H} = \gamma(\hat{b}_+^\dagger \hat{b}_+ + \hat{b}_-^\dagger \hat{b}_-) + \delta\hat{1}, \quad \gamma, \delta \in \mathbf{R}.$$

Evidently, the eigenstates of the number operators $b_\pm^\dagger \hat{b}_\pm$ are stationary states for this system. One says that the $(b^\dagger, b)$ operators create and destroy fermionic "quasi-particles" with energy γ. Show that the vacuum state of the original fermionic system contains pairs of quasi-particles.

Why this is an interesting problem: The transformation mixing creation and annihilation operators is named after Bogoliubov who used it to understand the BCS model of superconductivity. In this context the quasi-particles are "Cooper pairs" of electrons. A generalization of this model with a more complicated Bogoliubov transformation can be used to understand the "Schwinger effect", the "Unruh effect", and the "Hawking effect", which are three important quantum field theoretic phenomena you should read about.

Chapter 5

Electromagnetic Fields

The photon is a quantum manifestation of the electromagnetic field. To see what that means we will need to review some electrodynamics.

5.1 Maxwell equations

I will use the Gaussian system of electromagnetic units and I will let c denote the speed of light in vacuum. The Maxwell equations are differential equations for the electric field $\vec{E}(\vec{x},t)$, and the magnetic field $\vec{B}(\vec{x},t)$, which are defined by the force they exert on a test charge q at the point $\vec{x}$ at time t via the *Lorentz force law*:

$$\vec{F}(\vec{x},t) = q\left(\vec{E}(\vec{x},t) + \frac{1}{c}\vec{v}(t)\times\vec{B}(\vec{x},t)\right), \tag{5.1}$$

where $\vec{v}(t)$ is the particle's velocity at time t. Equation (5.1) is used to determine the motion of a charged particle in a given electromagnetic field assuming the effect of the particle on the field can be neglected. Equation (5.1) can also be used to measure the electromagnetic field by observing the motion of charged "test" particles.

The Lorentz force law tells us how the electromagnetic field affects electrically charged matter. The Maxwell equations tell us how the electrically charged matter affects the electromagnetic field. In macroscopic applications it is usually convenient to model the electric charges – the "sources" – of the electromagnetic field as a continuous electric charge density $\rho(\vec{x},t)$ and electric current density $\vec{j}(\vec{x},t)$. The Maxwell equations are

$$\nabla\cdot\vec{E} = 4\pi\rho, \tag{5.2}$$

$$\nabla\cdot\vec{B} = 0, \tag{5.3}$$

$$\nabla\times\vec{B} - \frac{1}{c}\frac{\partial\vec{E}}{\partial t} = \frac{4\pi}{c}\vec{j}, \tag{5.4}$$

$$\nabla\times\vec{E} + \frac{1}{c}\frac{\partial\vec{B}}{\partial t} = 0. \tag{5.5}$$

I would like to spend a little time getting a feel for the basic features of these equations.

5.1.1 The basic structure of the Maxwell equations

First of all, the equations (5.2)–(5.5) are 8 coupled, first-order, partial differential equations (with constant coefficients in Cartesian coordinates) for the 6 unknown functions contained in the components of $\vec{E}$ and $\vec{B}$.[1] One usually views the Maxwell equations as equations which are used to determine $\vec{E}$ and $\vec{B}$ for a given ρ and $\vec{j}$. In this setting the equations for $\vec{E}$ and $\vec{B}$ are "linear-inhomogeneous" thanks to the "source terms" defined by ρ and $\vec{j}$. Because the equations are inhomogeneous, it is not possible to superimpose solutions $(\vec{E}_1, \vec{B}_1)$ and $(\vec{E}_2, \vec{B}_2)$ to get new solutions without altering the charge and current densities. On the other hand, given any solution to the Maxwell equations (for a given ρ and $\vec{j}$) one can add any solution of the source-free Maxwell equations (obtained by setting $\rho = 0 = \vec{j}$) to get a new solution of the same inhomogeneous equations. If one is solving the Maxwell equations in a region of space where there are no sources, $\rho = 0$ and $\vec{j} = 0$, then the equations are homogeneous and one *can* always superimpose solutions.

The equations (5.2) and (5.3) represent 2 "scalar" equations, while equations (5.4) and (5.5) are "vector equations". A vector equation equates the components of two vectors. Thus the equations (5.4) and (5.5) each represent 3 (coupled) equations in which the x component of the left-hand side is equated to the x component of the right hand side, and so on.

As already mentioned, the Maxwell equations are typically solved for $\vec{E}$ and $\vec{B}$ once the charge density and its motion (the current density) are specified. For example, one can let the charge density be that of a uniform ball of positive charge held fixed in space so that the current density vanishes. As you might guess, the solution of these equations has vanishing magnetic field and a Coulomb-type electrostatic field outside the ball. (Do you remember what happens inside the ball?) Note that this way of using the Maxwell equations assumes that the motion of the sources is completely known (or else, how could we specify ρ and $\vec{j}$?). For many purposes this is a reasonable physical assumption. But, strictly speaking, this is only an approximate description of electrodynamics. As you can imagine, many applications (*e.g.,* the electrodynamics of the ionosphere) will require us to also figure out how the sources are moving. This is a complicated problem and quite non-linear: the sources generate the electromagnetic field according to the Maxwell equations (5.2)–(5.5), the electromagnetic field affects the sources according to the Lorentz force law (5.1), but the motion of the charges determines the fields, and so on. Of course, if there are no sources this is not an issue – at least classically. It turns out that in the quantum domain the sources can never be completely ignored, although we won't develop the theory far enough to see that.

Note that only 4 of the 8 Maxwell equations, (5.2) and (5.4), involve the sources. These are often called the *inhomogeneous equations* because they are linear inhomogeneous in the unknowns. The other 4 which do not involve the sources, (5.3) and (5.5), are likewise known as the *homogeneous equations*. The inhomogeneous equation (5.2) involving ρ – often called the *Gauss law* – shows that the charge density gives a divergence to the electric field. This is one way in which electric charges create electric fields. The Maxwell equation (5.3), on the other hand, shows that the magnetic field never has a divergence. By analogy with the electric field, this equation can be viewed as stating that there is no such thing as "magnetic charge". The magnetic field can have a curl, however. Eq. (5.4) asserts that this corresponds to having a time varying electric field and/or having an electric current density (moving charges). Thus a moving charge creates a magnetic field (the *Biot-Savart law*) and, as Maxwell first postulated, so does a time varying electric field. Finally, from (5.5), the electric field can also have a curl, but only if there is a time varying magnetic field—a phenomenon

[1] Just by counting equations and unknowns one might worry that the system of equation is over-determined. We shall see that this is not the case thanks to some fundamental identities satisfied by the equations.

characterized by Faraday and one which is necessary for the existence of civilization as we know it.

Note also that only 6 of the 8 equations, (5.4) and (5.5), involve a time derivative, that is, only 6 equations concern themselves with how the fields change in time. For this reason these equations are called the *evolution equations*. The remaining two divergence equations are called *constraint equations* since they restrict the fields at any given time. It can be shown that the constraint equations only need to be satisfied once, *i.e.,* at a single instant of time; the evolution equations will guarantee they will be satisfied at later times. This is important because otherwise we would be in danger of having too many equations (8) and not enough unknowns (6).

5.1.2 Continuity equation and conservation of electric charge

I would now like to consider an important consistency condition that must be satisfied by the sources if the Maxwell equations are to admit any solution at all. This consistency condition says that a necessary condition for the existence of a solution $(\vec{E}, \vec{B})$ to the Maxwell equations is that the sources $(\rho, \vec{j})$ must satisfy a continuity equation.

Take the time derivative of (5.2) and interchange time and space derivatives to get

$$\nabla \cdot \frac{\partial \vec{E}}{\partial t} = 4\pi \frac{\partial \rho}{\partial t}.$$

Compare this result with the divergence of (5.4) :

$$-\frac{1}{c}\nabla \cdot \frac{\partial \vec{E}}{\partial t} = \frac{4\pi}{c}\nabla \cdot \vec{j},$$

to find

$$\frac{\partial \rho}{\partial t} + \nabla \cdot \vec{j} = 0.$$

What this computation means is that the Maxwell equations have solutions for $\vec{E}$ and $\vec{B}$ only if the 4 functions $\rho(\vec{x}, t)$ and $\vec{j}(\vec{x}, t)$ are chosen to satisfy the continuity equation (5.1.2) given above. Recall that this equation is a differential version of a conservation law; the conserved quantity in this case being the electric charge. Explicitly, the total charge Q contained in a volume V at time t, defined by

$$Q = \int_V dV\, \rho(\vec{x}, t),$$

changes in time according to the net flux of the current density $\vec{j}$ through the boundary S of V:

$$\frac{dQ}{dt} = -\oint_S d\vec{S} \cdot \vec{j}.$$

If the net flux of charge through the boundary (which may be "at infinity") vanishes, then the charge contained in V is constant in time. When we use the Maxwell equations to solve for the electromagnetic field due to a given charge distribution, that distribution must be specified so that charge is conserved in the sense of (5.1.2) or else the equations cannot have a solution.[2]

[2]It is no accident that the Maxwell equations, in effect, force the conservation of electric charge. Indeed, our current field theoretic description of all fundamental interactions (electromagnetic, weak, strong, and gravitational) is geared to force such conservation laws through the use of variational principles and the "principle of local gauge invariance". Unfortunately, a discussion of such ideas would be beyond the scope of this course.

Given the continuity equation, we can now consider the status of the constraint equations (5.2) and (5.3). It is straightforward to show that if they are satisfied at one time, say $t = 0$, by the initial values for $\vec{E}$ and $\vec{B}$, then they are automatically solved at later times provided (i) the electromagnetic field at later times satisfies the evolution equations, and (ii) (5.1.2) is satisfied by the sources. See the Problems for details.

5.2 The electromagnetic wave equation

Let us now see how the Maxwell equations (5.2)–(5.5) predict the existence of electromagnetic waves. In all that follows I will consider a region of space and time in which there are no sources (*i.e.,* we will consider the propagation of electromagnetic waves in vacuum). We shall see that there are wave solutions to the Maxwell equations in source-free regions. Indeed, the general solution to the Maxwell equations in vacuum will be a superposition of these wave solutions. These "radiative degrees of freedom" of the electromagnetic field will turn out to be what we need to model the phenomena of photons.

In what follows I will set $\rho = 0 = \vec{j}$ in our space-time region of interest. Now all the Maxwell equations are linear, homogeneous.

Begin by taking the curl of (5.4) to find

$$\nabla\times(\nabla\times\vec{B}) - \frac{1}{c}\nabla\times\frac{\partial\vec{E}}{\partial t} = 0. \tag{5.6}$$

Of course we assume that the components of the electromagnetic field are at least twice continuously differentiable functions of space and time. We are therefore allowed to assume that partial derivatives commute and we have, using the identity[3]

$$\nabla \times \nabla \times \vec{A} = \nabla(\nabla \cdot \vec{A}) - \nabla^2\vec{A} \tag{5.7}$$

$$\nabla(\nabla\cdot\vec{B}) - \nabla^2\vec{B} - \frac{1}{c}\frac{\partial}{\partial t}\nabla\times\vec{E} = 0. \tag{5.8}$$

Next we use the fact that (i) the magnetic field is divergence-free (see (5.3)) and (ii) the equation (5.5) to find

$$-\nabla^2\vec{B} + \frac{1}{c^2}\frac{\partial^2\vec{B}}{\partial t^2} = 0. \tag{5.9}$$

This says that each component of $\vec{B}$ satisfies the wave equation with wave velocity c! You can play an analogous game with the electric field to show

$$-\nabla^2\vec{E} + \frac{1}{c^2}\frac{\partial^2\vec{E}}{\partial t^2} = 0. \tag{5.10}$$

So, each component of $\vec{E}$ and $\vec{B}$ satisfies a wave equation, *i.e.,* we can have traveling wave disturbances in the electromagnetic field. As you know, such disturbances correspond to light waves, radio waves, microwaves, and so forth.

[3]The vector Laplacian requires some definition in curvilinear coordinates. The Laplacian of the vector is just the scalar Laplacian of the components in Cartesian coordinates.

From our computations above it is tempting to think that the Maxwell equations are really just 6 copies of the wave equation. In particular it appears that each component of the electric and magnetic fields goes its own way, independently of the others. This is not true. What we *have* shown is that, given an $\vec{E}$ and $\vec{B}$ solving the (source-free) Maxwell equations, the electric and magnetic fields necessarily satisfy uncoupled wave equations. What we have *not* shown — and is not true — is that, given electric and magnetic fields solving uncoupled wave equations (5.9) and (5.10), we obtain solutions to the Maxwell equations. Put differently, that $\vec{E}$ and $\vec{B}$ solve the wave equations above is *necessary but not sufficient* for them to solve the source-free Maxwell equations. I will now show how to use the solutions to (5.9) and (5.10) to build *bona fide* solutions to the Maxwell equations in the form of electromagnetic waves. For simplicity, I will restrict attention to electromagnetic plane waves. (More general waves can be obtained by superposition in the usual way.) The strategy is to take a plane wave solution to (5.10), restrict it to satisfy (5.2) (with $\rho = 0$) and then define the magnetic field so that the remaining Maxwell equations are satisfied. Here are the details.

Because each component of $\vec{E}$ satisfies the wave equation, we can consider a plane wave electric field of the form:

$$\vec{E}(\vec{x}, t) = \vec{E}_0 \cos(\vec{k} \cdot \vec{x} - \omega t + \phi). \tag{5.11}$$

Here the constant vector field $\vec{E}_0$ determines the amplitude and (constant) direction of the electric field, and the phase ϕ is a constant. You can check that (5.11) is a solution to (5.10) if and only if the dispersion relation is

$$\omega = kc. \tag{5.12}$$

Of course we could also use a sine function, or take the real part of an exponential, or take linear combinations (the source-free Maxwell equations are linear homogeneous), but this is good enough for our discussion. I emphasize that while each component of (5.11) solves the wave equation when (5.12) holds, (5.11) does not yet necessarily define a solution to the Maxwell equations. The Maxwell equations still need to be imposed. To begin with, let us impose $\nabla \cdot \vec{E} = 0$. You can easily check that this equation means we have to choose the wave vector to be orthogonal to the amplitude vector:

$$\vec{k} \cdot \vec{E}_0 = 0. \tag{5.13}$$

Thus the direction of the electric field in the electromagnetic plane wave is always perpendicular to the direction of propagation of the wave. As you probably know, electromagnetic waves are "transverse".

The remaining Maxwell equations still need to be imposed, and they involve the magnetic field. We can build a divergence-free ($\nabla \cdot \vec{B} = 0$) magnetic field plane wave just as we did for $\vec{E}$, but we must keep in mind the other 2 Maxwell equations:

$$\nabla \times \vec{B} - \frac{1}{c} \frac{\partial \vec{E}}{\partial t} = 0 \tag{5.14}$$

$$\nabla \times \vec{E} + \frac{1}{c} \frac{\partial \vec{B}}{\partial t} = 0. \tag{5.15}$$

Let us first look at (5.14). From (5.11) the time derivative of $\vec{E}$ yields a sine function, and this is to be equated to a curl of $\vec{B}$, which involves first spatial derivatives. This suggests we should also try a cosine for the magnetic field:

$$\vec{B} = \vec{B}_0 \cos(\vec{k}' \cdot \vec{x} - \omega' t + \psi). \tag{5.16}$$

To satisfy the wave equation we need the dispersion relation

$$\omega' = k'c. \tag{5.17}$$

Because we want $\nabla \cdot \vec{B} = 0$, we have to choose the wave vector and amplitude such that

$$\vec{k}' \cdot \vec{B}_0 = 0. \tag{5.18}$$

It is easy to check that

$$\nabla \times \vec{B} = -\vec{B}_0 \times \nabla \left[\cos(\vec{k}' \cdot \vec{x} - \omega' t + \psi)\right] \tag{5.19}$$

$$= (\vec{B}_0 \times \vec{k}') \sin(\vec{k}' \cdot \vec{x} - \omega' t + \psi). \tag{5.20}$$

Next, we compute

$$\frac{1}{c}\frac{\partial \vec{E}}{\partial t} = \frac{\omega}{c}\vec{E}_0 \sin(\vec{k} \cdot \vec{x} - \omega t + \phi). \tag{5.21}$$

Thus (5.14) reduces to

$$\frac{\omega}{c}\vec{E}_0 \sin(\vec{k} \cdot \vec{x} - \omega t + \phi) = \vec{B}_0 \times \vec{k}' \sin(\vec{k}' \cdot \vec{x} - \omega' t + \psi). \tag{5.22}$$

Since this must hold at each location and time, it follows that

$$\phi = \psi, \tag{5.23}$$

$$\vec{k}' = \vec{k}, \tag{5.24}$$

and

$$\frac{\omega}{c}\vec{E}_0 = -\vec{k} \times \vec{B}_0. \tag{5.25}$$

Let $\hat{n}$ be the unit vector in the direction of wave propagation:

$$\hat{n} = \frac{\vec{k}}{k}, \tag{5.26}$$

then we can write

$$\vec{E}_0 = -\hat{n} \times \vec{B}_0. \tag{5.27}$$

Thus we see that the electric and magnetic fields must have the same wave vector, that is, they have the same wavelengths ($\lambda = 2\pi/k$) and propagation directions $\frac{\vec{k}}{k}$. Furthermore, from (5.27) and the dispersion relation (5.12) it follows that:

- (1) The magnitudes of the electric and magnetic fields are equal[4],

$$E_0 = B_0.$$

- (2) $\vec{E}$ and $\vec{B}$ are orthogonal, $\vec{E} \cdot \vec{B} = 0$.
- (3) Each field is orthogonal to the propagation direction, $\vec{E} \cdot \hat{n} = 0 = \vec{B} \cdot \hat{n}$.

[4]Keep in mind we are using Gaussian units.

Note that when $\vec{E}_0$ and $\vec{B}_0$ satisfy (5.27), they automatically satisfy (5.13) and (5.18).

Thus far we have only dealt with (5.14). As an exercise you can show the other evolution equation (5.15) gives identical information

At this point we have taken care of all the Maxwell equations. Let us gather all our results together. We can build a plane wave solution to the (source-free) Maxwell equations by the following procedure:

- (i) Pick a wave propagation direction $\hat{n}$ and a wavelength λ; then define

$$\vec{k} = \frac{2\pi}{\lambda}\hat{n}.$$

- (ii) Pick an electric field amplitude $\vec{E}_0$, which is any (constant) vector (field) such that $\vec{E}_0 \cdot \hat{n} = 0$.

- (iii) Set $\omega = kc$ and define

$$\vec{E}(\vec{x}, t) = \vec{E}_0 \cos(\vec{k} \cdot \vec{x} - \omega t + \phi), \tag{5.28}$$

$$\vec{B}(\vec{x}, t) = (\hat{n} \times \vec{E}_0) \cos(\vec{k} \cdot \vec{x} - \omega t + \phi). \tag{5.29}$$

Equations (5.28) and (5.29) give the electric and magnetic fields arising in a plane electromagnetic wave. This wave is *linearly polarized*. This means that the electric field always points in the same direction. Later we shall consider linear combinations in which the electric field direction traces out a circle as the wave travels. Such solutions are *circularly polarized*.

It can be shown that the general solution to the source-free Maxwell equations can be expressed as a superposition of these plane wave solutions over all possible values of the vector amplitude, phase, propagation direction, and wave number.

5.3 Electromagnetic energy, momentum, and angular momentum

In a previous physics course you should have encountered the interesting notion that the electromagnetic field carries energy and momentum. If you have ever been sunburned you have experimental confirmation of this fact! In physics, the notions of energy, momentum, and angular momentum are of particular interest because they are conserved quantities under appropriate circumstances. The electromagnetic field manifests each of these conserved quantities and they have an important role to play in the physical properties of the photon.

We can uncover the energy, momentum, and angular momentum of the electromagnetic fields by searching for conservation laws. Such conservation laws will appear embodied in continuity equations, which are derived from the Maxwell equations. I will show you how the energy conservation works in detail and just state the results for momentum and angular momentum.

5.3.1 Energy

Let us begin by checking the conservation of energy. I will restrict attention to the source-free case ($\rho = 0$, $\vec{j} = 0$). With sources, there can be an exchange of energy-momentum of the electromagnetic field with that of the sources and the story is a little longer than justified for our needs[5]. Since we have used the symbols ρ and $\vec{j}$ to denote the electric charge density and electric current density, to avoid confusion I will use the symbols u and $\vec{s}$ to denote the energy density and energy current density (also known as the "Poynting vector") for the electromagnetic field. These are defined by (still in Gaussian units)

$$u = \frac{1}{8\pi}(E^2 + B^2), \tag{5.30}$$

$$\vec{s} = \frac{c}{4\pi}\vec{E} \times \vec{B}. \tag{5.31}$$

The claim is that $(u, \vec{s})$ satisfy

$$\frac{\partial u}{\partial t} + \nabla{\cdot}\vec{s} = 0 \tag{5.32}$$

when $\vec{E}$ and $\vec{B}$ satisfy the (source-free) Maxwell equations.

To verify this, we need to use some vector identities. If $\vec{A}$ and $\vec{C}$ are any two vector fields that depend upon a variable t, we have that

$$\frac{\partial}{\partial t}(A^2) = \frac{\partial}{\partial t}(\vec{A} \cdot \vec{A}) \tag{5.33}$$

$$= 2\vec{A} \cdot \frac{\partial \vec{A}}{\partial t}, \tag{5.34}$$

and

$$\nabla{\cdot}(\vec{A} \times \vec{C}) = \vec{C} \cdot \nabla{\times}\vec{A} - \vec{A} \cdot \nabla{\times}\vec{C}. \tag{5.35}$$

Using these vector identities and the Maxwell evolution equations, we have

$$\frac{\partial u}{\partial t} = \frac{1}{4\pi}\left(\vec{E} \cdot \frac{\partial \vec{E}}{\partial t} + \vec{B} \cdot \frac{\partial \vec{B}}{\partial t}\right) \tag{5.36}$$

$$= \frac{c}{4\pi}\left(\vec{E} \cdot \nabla{\times}\vec{B} - c\vec{B} \cdot \nabla{\times}\vec{E}\right), \tag{5.37}$$

and

$$\nabla{\cdot}\vec{s} = \frac{c}{4\pi}\left(\vec{B} \cdot \nabla{\times}\vec{E} - \vec{E} \cdot \nabla{\times}\vec{B}\right), \tag{5.38}$$

from which the result (5.32) follows immediately. Note we only needed the evolution equations to obtain (5.32).

We interpret $u(\vec{x}, t)$ as the *electromagnetic energy density at the point $\vec{x}$ and time t* and the Poynting vector[6] $\vec{s}$ as the energy current density. Define the total electromagnetic energy U in the volume V to be the integral of the energy density u:

$$U := \int_V dV\, u = \frac{1}{8\pi}\int_V dV\, \left(E^2 + B^2\right). \tag{5.39}$$

[5] As you might expect on physical grounds, in order to get a conservation law in this case, one must keep track of the energy and momentum of the sources.

[6] You should keep in mind that the Poynting vector is really a vector *field*.

According to the continuity equation (5.32), the time rate of change of U is controlled by the flux of the Poynting vector $\vec{s}$ through the boundary A of V:

$$\frac{dU}{dt} = -\int_A d\vec{A} \cdot \vec{s}, \tag{5.40}$$

where

$$d\vec{A} = \hat{N}\, dA \tag{5.41}$$

with $\hat{N}$ being the outward unit normal to the boundary surface of V.

It is worth computing the energy density and Poynting vector for the plane wave constructed above in (5.28) and (5.29). You will find

$$u = \frac{1}{4\pi}(E_0)^2 \cos^2(\vec{k} \cdot \vec{x} - \omega t + \phi), \tag{5.42}$$

$$\vec{s} = \frac{c}{4\pi}(E_0)^2 \cos^2(\vec{k} \cdot \vec{x} - \omega t + \phi)\, \hat{n}. \tag{5.43}$$

We see that the energy density of the wave is proportional to the amplitude-squared of the electric field. We also see that that the Poynting vector lies along the direction of propagation $\hat{n} = \vec{k}/k$ of the plane wave. Thus the flow of energy is along $\hat{n}$. The continuity equation guarantees that if you integrate u over a volume you will find that the total energy is changing according to the net flux of $\vec{s}$ through the volume.

If, at the boundary of the region V, the fields are such that net the Poynting flux vanishes, then the energy contained in V will be constant in time. In particular, if the volume V in (5.39) is taken to be all space, we can view the boundary A as a sphere of infinite radius. If we consider an *isolated system*, so that the electric and magnetic fields vanish sufficiently rapidly at large distances (*i.e.,* "at infinity"), then the flux of the Poynting vector will vanish as the radius of A is taken to infinity. Thus the total electromagnetic energy of an isolated (and source-free) electromagnetic field is constant in time.

5.3.2 Momentum and angular momentum

Next let's briefly look at electromagnetic momentum and angular momentum and their conservation laws. It turns out that the momentum density $\vec{\pi}$ for the electromagnetic field is proportional to the Poynting vector:

$$\vec{\pi} = \frac{1}{c^2}\vec{s} = \frac{1}{4\pi c}\vec{E} \times \vec{B}. \tag{5.44}$$

Using the Maxwell equations it can be shown that each Cartesian component π^i of $\vec{\pi}$ satisfies

$$\frac{\partial}{\partial t}\pi^i + \sum_{j=1}^{3} \frac{\partial}{\partial x^j}(\tau^{ij}) = 0, \tag{5.45}$$

where τ^{ij} are the components of a tensor which is minus the *Maxwell stress tensor*. It is not too hard to show this – you could try it, but I will leave a proof of this to a course in electrodynamics. It is perhaps intuitively clear that the Poynting vector, which defines the transport of energy should also characterize the momentum. Each Cartesian component of $\vec{\pi}$ is a momentum density which

can be integrated to get a conserved quantity. The volume integral of each Cartesian component of $\vec{\pi}$ represents a conserved momentum $\vec{P}$ for the electromagnetic field:

$$\vec{P} = \int_V dV \frac{1}{4\pi c} \vec{E} \times \vec{B}. \tag{5.46}$$

The *angular momentum density* for the electromagnetic field is

$$\vec{l} = \vec{x} \times \vec{\pi} = \frac{1}{4\pi c} \vec{x} \times (\vec{E} \times \vec{B}). \tag{5.47}$$

Notice that, as usual, the notion of angular momentum depends upon a reference point – here used as a choice of origin – from which to define the position vector $\vec{x}$. Each component of the angular momentum density satisfies a continuity equation when the (source-free) Maxwell equations hold. The total angular momentum in a region V is

$$\vec{L} = \frac{1}{4\pi c} \int_V dV \left[\vec{x} \times (\vec{E} \times \vec{B}) \right], \tag{5.48}$$

where it must be understood that this formula is meant to be applied to the Cartesian components of the various vectors.

As you know, the Maxwell equations define a field theory in accord with the principles of special relativity. In this context, I should note that the energy, momentum, and angular momentum have been defined in a given inertial reference frame. Under a change of reference frame the energy and momentum mix; the reference frame independent quantity is the *energy-momentum 4-vector* $\mathcal{P} = (U/c, \vec{P})$. The behavior of the angular momentum under a change of inertial reference frame is somewhat more complicated; it is properly viewed as part of a relativistic angular momentum tensor. See, *e.g.,* [5] for details.

5.4 Electromagnetic potentials

The homogeneous subset of the Maxwell equations,

$$\nabla \times \vec{E} + \frac{1}{c} \frac{\partial \vec{B}}{\partial t} = 0 \tag{5.49}$$

and

$$\nabla \cdot \vec{B} = 0, \tag{5.50}$$

are equivalent to the existence of a vector field, the *vector potential* $\vec{A}$, and a scalar field, the *scalar potential* ϕ, such that

$$\vec{E} = -\frac{1}{c} \frac{\partial \vec{A}}{\partial t} - \nabla \phi, \quad \vec{B} = \nabla \times \vec{A}. \tag{5.51}$$

With $(\phi, \vec{A})$ arbitrary, this is the general solution to the half of Maxwell's equations given by (5.49), (5.50).[7] Once we introduce the potentials the only remaining Maxwell equations to be solved are the inhomogeneous ones (5.2), (5.4).

[7] Technically, this is the general solution "in the small"; boundary conditions may put additional restrictions on the potentials.

The potentials are far from uniquely determined by $\vec{E}$ and $\vec{B}$. If $(\phi, \vec{A})$ are a set of potentials which yield a given electromagnetic field $(\vec{E}, \vec{B})$ via (5.51), then so are (exercise)

$$\phi' = \phi - \frac{1}{c}\frac{\partial \Lambda}{\partial t}, \tag{5.52}$$

$$\vec{A}' = \vec{A} + \nabla \Lambda, \tag{5.53}$$

where Λ is *any* function of space and time. You can easily check this; just substitute (5.52), (5.53) into (5.51) and watch the terms involving Λ drop out. This transformation between two sets of potentials for the same electromagnetic field is called a *gauge transformation* for historical reasons that I shall not go into. Under a gauge transformation the electromagnetic field is unchanged; we say the electric and magnetic fields are *gauge invariant*. The notions of gauge transformations and gauge invariance, which may just seem to you like a technical detail in Maxwell theory, is actually pretty profound in physics. For our purposes, the freedom to redefine potentials via a gauge transformation means that we have some freedom to put the potentials into a convenient form by making a "choice of gauge".

For our purposes we will use potentials which are in the *radiation gauge*. What this means is as follows. Any solution to the Maxwell equation $(\vec{E}, \vec{B})$ *in the absence of sources*, $\rho = 0 = \vec{j}$, can be described by a set of potentials such that

$$\phi = 0, \quad \nabla \cdot \vec{A} = 0. \tag{5.54}$$

To see this, given any potentials ϕ_1 and $\vec{A}_1$ for the given $(\vec{E}, \vec{B})$, we can make a gauge transformation using a function Λ_1 which at each time t is a solution to the Poisson equation

$$\nabla^2 \Lambda_1 = -\nabla \cdot \vec{A}_1. \tag{5.55}$$

The result is a set of potentials $(\phi_2, \vec{A}_2)$ where

$$\nabla \cdot \vec{A}_2 = 0. \tag{5.56}$$

Such potentials are said to be in the *Coulomb gauge*. (The Coulomb gauge is available whether or not the sources vanish.) One can make a further gauge transformation by a function Λ_2, preserving the Coulomb gauge, provided

$$\nabla^2 \Lambda_2 = 0. \tag{5.57}$$

To gauge transform the scalar potential to zero one should choose

$$\Lambda_2 = c \int dt\, \phi_2. \tag{5.58}$$

But this choice will satisfy (5.57) only if $\nabla^2 \phi_2 = 0$. If there are sources this need not be the case because in the Coulomb gauge

$$\nabla \cdot \vec{E} = 4\pi\rho \quad \Longleftrightarrow \quad \nabla^2 \phi = -4\pi\rho. \tag{5.59}$$

But in a region where $\rho = 0$ there is no obstruction to transforming the scalar potential to zero.

In the radiation gauge the electromagnetic fields are determined by the electromagnetic potentials according to[8]

$$\vec{E} = -\frac{1}{c}\frac{\partial \vec{A}}{\partial t}, \quad \vec{B} = \nabla \times \vec{A}. \tag{5.60}$$

Any gauge transformation Λ which preserves the radiation gauge must satisfy

$$\frac{\partial \Lambda}{\partial t} = 0, \quad \nabla^2 \Lambda = 0. \tag{5.61}$$

With suitable boundary conditions one can ensure that $\Lambda = const.$ is the only solution. Such gauge transformations are trivial – they do not transform anything and we identify them with the identity transformation. In this way the potentials can be uniquely associated to the electromagnetic field in the radiation gauge.

By using potentials half the Maxwell equations are satisfied. The remaining equations are (5.2) and (5.4) with $\rho = 0 = \vec{j}$,

$$\nabla \cdot \vec{E} = 0, \tag{5.62}$$

and

$$\nabla \times \vec{B} - \frac{1}{c}\frac{\partial \vec{E}}{\partial t} = 0 \tag{5.63}$$

Substituting for $(\vec{E}, \vec{B})$ using (5.60 shows that (5.62) is satisfied in the radiation gauge and (5.63) becomes

$$\frac{1}{c^2}\frac{\partial^2 \vec{A}}{\partial t^2} - \nabla^2 \vec{A} = 0, \quad \nabla \cdot \vec{A} = 0 \tag{5.64}$$

where I used the vector identity (5.7). Thus the entire set of source-free Maxwell equations can be expressed as the wave equation (5.64) for a "transverse" vector field via the radiation gauge.

5.5 Role of the sources

Let me briefly comment on how the sources enter into this game. Allowing for the presence of sources means we cannot set $\phi = 0$, but we can still use the Coulomb gauge $\nabla \cdot \vec{A} = 0$. In the Coulomb gauge the scalar potential satisfies the Poisson equation (5.59). With appropriate boundary conditions this will uniquely determine the scalar potential in terms of the charge density. Having eliminated the scalar potential in this way, the electromagnetic degrees of freedom will (still) reside in the transverse vector potential, which now satisfies

$$\frac{1}{c^2}\frac{\partial^2 \vec{A}}{\partial t^2} - \nabla^2 \vec{A} = \frac{4\pi}{c}\vec{j}_T, \quad \nabla \cdot \vec{A} = 0 = \nabla \cdot \vec{j}_T. \tag{5.65}$$

where $\vec{j}_T$ is the *transverse electric current*:

$$\vec{j}_T = \vec{j} - \frac{1}{4\pi c}\frac{\partial}{\partial t}\nabla\phi. \tag{5.66}$$

[8] This might amuse you (at least a little). In electrostatics we have $\vec{B} = 0$ and $\nabla \times \vec{E} = 0$; it is conventional to work in a gauge in which $\vec{A} = 0$ so that the static electric field is given by $\vec{E} = -\nabla\phi$. This is certainly the most convenient way to analyze electrostatics, but one *could* opt to set the scalar potential to zero and use a time-dependent (and curl-free) vector potential if so-desired.

As you may know from a class in electrodynamics, the solution of these equations can be obtained as

$$\vec{A} = \vec{A}_{\vec{j}} + \vec{A}_0, \tag{5.67}$$

where $\vec{A}_{\vec{j}}$ is determined by $\vec{j}_T$ using the retarded Green function and $\vec{A}_0$ is any solution to the source-free equations. In this sense we say that the solutions to the source-free equations represent the "degrees of freedom" of the electromagnetic field. It is these degrees of freedom which become interpreted in terms of photons in the quantum theory.

As a simple example of all this, suppose the source consists of a point charge q at rest at the origin (in a given inertial reference frame) so that

$$\rho(\vec{x}, t) = q\delta(\vec{x}), \quad \vec{j}(\vec{x}, t) = \vec{j}_T(\vec{x}, t) = 0. \tag{5.68}$$

Let the spatial region of interest be all of Euclidean space, with fields vanishing "at infinity". The scalar potential is then the usual Coulomb field, $\phi = q/|\vec{x}|$, and we have $\vec{A}_{\vec{j}} = 0$. The homogeneous solutions $\vec{A}_0$ are unrestricted (except by the asymptotic conditions). The presence of $\vec{A}_0$ represents the presence of electromagnetic radiation in addition to the Coulomb field. All together, this is the classical model of the electromagnetic field experienced by an electron in a hydrogenic atom in the presence of radiation. In Chapter 7 we shall use the quantum version of $\vec{A}_0$ to understand spontaneous emission.

5.6 Solution to the source-free Maxwell equations

The general solution to (5.64) can be obtained in a useful form from some Fourier analysis. Begin by making a Fourier decomposition of $\vec{A}$ at each time t:

$$\vec{A}(\vec{x}, t) = \frac{1}{(2\pi)^{3/2}} \int_{\mathbf{R}^3} d^3k \, \vec{\mathcal{A}}(\vec{k}, t) e^{i\vec{k}\cdot\vec{x}}, \tag{5.69}$$

$$\vec{\mathcal{A}}(\vec{k}, t) = \frac{1}{(2\pi)^{3/2}} \int_{\text{all space}} d^3x \, \vec{A}(\vec{x}, t) e^{-i\vec{k}\cdot\vec{x}}. \tag{5.70}$$

For notational simplicity I am denoting integrals over Fourier space by

$$\int_{\mathbf{R}^3} d^3k \equiv \int_{-\infty}^{\infty} dk_x \int_{-\infty}^{\infty} dk_y \int_{-\infty}^{\infty} dk_z. \tag{5.71}$$

At each time t the Fourier transform $\vec{\mathcal{A}}(\vec{k}, t)$ is a vector of complex functions of the real variables $\vec{k} = (k_x, k_y, k_z)$. Because $\vec{A}(\vec{x}, t)$ is real its transform must satisfy

$$\vec{\mathcal{A}}(\vec{k}, t)^* = \vec{\mathcal{A}}(-\vec{k}, t). \tag{5.72}$$

The gauge condition which $\vec{A}$ satisfies implies that its Fourier transform is orthogonal to $\vec{k}$:

$$0 = \nabla \cdot \vec{A} = \frac{1}{(2\pi)^{3/2}} \int_{\mathbf{R}^3} d^3k \, (i\vec{k}) \cdot \vec{\mathcal{A}}(\vec{k}, t) e^{i\vec{k}\cdot\vec{x}} \quad \Longleftrightarrow \quad \vec{k} \cdot \vec{\mathcal{A}}(\vec{k}, t) = 0. \tag{5.73}$$

In terms of the Fourier transform the electric and magnetic fields take the form:

$$\vec{E} = -\frac{1}{c}\frac{\partial \vec{A}}{\partial t} = -\frac{1}{(2\pi)^{3/2}}\int_{\mathbf{R}^3} d^3k\, \frac{1}{c}\frac{\partial}{\partial t}\vec{\mathcal{A}}(\vec{k},t)e^{i\vec{k}\cdot\vec{x}}, \tag{5.74}$$

$$\vec{B} = \nabla \times \vec{A} = \frac{i}{(2\pi)^{3/2}}\int_{\mathbf{R}^3} d^3k\, [\vec{k}\times\vec{\mathcal{A}}(\vec{k},t)]e^{i\vec{k}\cdot\vec{x}}. \tag{5.75}$$

Substitution of the Fourier expansion (5.69) into (5.64) shows that these equations are equivalent to

$$\ddot{\vec{\mathcal{A}}}(\vec{k},t) = -\omega^2(k)\vec{\mathcal{A}}(\vec{k},t), \quad \vec{k}\cdot\vec{\mathcal{A}}(\vec{k},t) = 0, \tag{5.76}$$

where

$$\omega(k) = kc. \tag{5.77}$$

This is pretty cool. For each $\vec{k}$, there are two complex components orthogonal to $\vec{k}$ in the Fourier component $\vec{A}(\vec{k},t)$. The above equations show that the real and imaginary parts of each independent component behaves like a harmonic oscillator with frequency $\omega(k)$!

Heuristically, you can think of the Fourier transform (5.70) as a generalization to a continuous infinity of degrees of freedom of the linear change of variables (2.21) to the displacements of the normal modes of vibration. In this generalization, for a given $\vec{k}$ (the real and imaginary part of) each component of $\vec{A}(\vec{k},t)$ orthogonal to $\vec{k}$ defines a normal mode of vibration. I will spare you some suspense and reveal that, in what follows, we will construct the quantum theory of these oscillators $\vec{A}(\vec{k},t)$ and identify the states of the system of oscillators with photon states. In this way a photon is built from quantum normal modes of vibration of the electromagnetic field.

With the harmonic oscillator interpretation in mind, we can write the solution of (5.76) in terms of complex amplitudes of vibration as follows:

$$\vec{\mathcal{A}}(\vec{k},t) = \sum_{\sigma=1}^{2}\sqrt{\frac{2\pi\hbar c}{k}}[a_\sigma(\vec{k})\,\vec{\varepsilon}_\sigma(\vec{k})\,e^{-i\omega(k)t} + a^*_\sigma(-\vec{k})\,\vec{\varepsilon}_\sigma(-\vec{k})\,e^{i\omega(k)t}]. \tag{5.78}$$

Here, for each $\vec{k}$, the vectors $\vec{\varepsilon}_\sigma(\vec{k})$, $\sigma = 1,2$ are two mutually orthogonal real-valued unit vectors, each orthogonal to $\vec{k}$,

$$\vec{\varepsilon}_\sigma(\vec{k})\cdot\vec{\varepsilon}_\tau(\vec{k}) = \delta_{\sigma\tau}, \quad \vec{\varepsilon}_\sigma(\vec{k})\cdot\vec{k} = 0, \quad \vec{\varepsilon}_1(\vec{k})\times\vec{\varepsilon}_2(\vec{k}) = \frac{\vec{k}}{k}. \tag{5.79}$$

The variables $a_\sigma(\vec{k})$ carry the information about the amplitude contribution for each wave with wave vector $\vec{k}$ as well as the polarization contribution. Note that we have used $\hbar$ to define the new variables. From a purely classical electromagnetic point of view this is a bit strange, but it is convenient for the quantum treatment we will give. For now, just think of the use of $\hbar$ as a way of giving convenient units to the amplitudes $a_\sigma(\vec{k})$, which are dimensionless (exercise).

We have now completely solved the source-free Maxwell equations in terms of a vector potential in the radiation gauge. To get the electric and magnetic fields one should substitute $\vec{A}$ as defined in (5.78)– (5.79) into (5.69) to get the vector potential $\vec{A}(\vec{x},t)$ in the radiation gauge:

$$\vec{A}(\vec{x},t) = \frac{1}{(2\pi)^{3/2}}\int_{\mathbf{R}^3} d^3k \sum_{\sigma=1}^{2}\sqrt{\frac{2\pi\hbar c}{k}}[a_\sigma(\vec{k})\,\vec{\varepsilon}_\sigma(\vec{k})\,e^{i\vec{k}\cdot\vec{x}-i\omega(k)t} + a^*_\sigma(\vec{k})\,\vec{\varepsilon}_\sigma(\vec{k})\,e^{-i\vec{k}\cdot\vec{x}+i\omega(k)t}]. \tag{5.80}$$

Substitute this potential into (5.60) to get $\vec{E}$ and $\vec{B}$ solving the source-free Maxwell equations. All solutions to the source-free Maxwell equations are obtained by making suitable choices for the complex amplitudes $a_\sigma(\vec{k})$.

5.7 Energy and momentum, revisited

In this section I will express the energy and momentum of the electromagnetic field in terms of the complex amplitudes for the harmonically oscillating normal modes derived in the previous section.

The energy for the electromagnetic field is given by

$$U = \frac{1}{8\pi}\int_{\text{all space}} d^3x\,(E^2 + B^2), \tag{5.81}$$

where it is understood the electric and magnetic fields are expressed in terms of the potentials via (5.51).[9] It is well worth knowing that in the Hamiltonian formalism for the electromagnetic field (see, *e.g.,* [5]) the quantity $-\vec{E}$ is ultimately identified with the momentum canonically conjugate to $\vec{A}$. You can then see from this energy formula that the E^2 term is then analogous to kinetic energy in Newtonian mechanics, while the B^2 term is analogous to potential energy. This perspective is underscored by the expression of the energy in terms of potentials. The "kinetic" and "potential" energies are

$$\frac{1}{8\pi}\int_{\text{all space}} d^3x\,E^2 = \frac{1}{8\pi}\int_{\text{all space}} d^3x\,\frac{1}{c^2}\left(\frac{\partial\vec{A}}{\partial t}\right)^2 \tag{5.82}$$

$$\frac{1}{8\pi}\int_{\text{all space}} d^3x\,B^2 = \frac{1}{8\pi}\int_{\text{all space}} d^3x\,\left(\nabla\times\vec{A}\right)^2. \tag{5.83}$$

Substituting in the Fourier expansion (5.69) and using (5.74), (5.75) yields

$$U = \frac{1}{8\pi c^2}\int_{\mathbf{R}^3} d^3k\left\{|\dot{\vec{A}}_{\vec{k}}|^2 + k^2c^2|A_{\vec{k}}|^2\right\}. \tag{5.84}$$

This form for U is of course consistent with our identification of the real and imaginary parts of the components of $\vec{A}_{\vec{k}}$ for each $\vec{k}$ as representing harmonic oscillators with frequency kc. In particular, the integral over Fourier space generalizes the sum over normal modes which would appear in the energy for a system of coupled oscillators.

It will be useful later to have the energy expressed in terms of the complex amplitudes. Using (5.78) and its consequence

$$\frac{1}{c}\dot{\vec{A}}_{\vec{k}} = -i\sqrt{2\pi\hbar kc}\sum_\sigma\left[a_\sigma(\vec{k})\epsilon_\sigma(\vec{k})e^{-i\omega t} - a^*_\sigma(-\vec{k})\epsilon_\sigma(-\vec{k})e^{i\omega t}\right], \tag{5.85}$$

we get, after just a little algebra,

$$U = \int_{\mathbf{R}^3} d^3k\sum_\sigma \hbar\omega(k)\,a^*_\sigma(\vec{k})a_\sigma(\vec{k}). \tag{5.86}$$

[9] Of course we will have restrict attention to fields which vanish at spatial infinity so that the integral exists.

You can use similar computations to get a nice formula for the momentum of the electromagnetic field. Recall that the momentum density for the electromagnetic field is given by (5.44). Substituting the Fourier form of the vector potential in terms of complex amplitudes as before, you will find

$$\vec{P} \equiv \int_{\text{all space}} d^3x \, \frac{1}{c} \vec{E} \times \vec{B} = \int d^3k \sum_\sigma \hbar \vec{k} \, a^*_\sigma(\vec{k}) a_\sigma(\vec{k}). \tag{5.87}$$

Hopefully, these very simple formulas for the energy and momentum justify all the effort that went into deriving them. Evidently, the electromagnetic field is a continuous family of coupled oscillators. The a and a^* variables are the complex amplitudes for the normal modes of vibration for the electromagnetic field. The fact that there are infinitely many of these oscillators is novel, and is needed to build up (by superposition) all the familiar – but not manifestly oscillator-like – electromagnetic phenomena. The formulas for the energy and momentum show that a normal mode of vibration with wave vector $\vec{k}$ and polarization σ is characterized by a dimensionless intensity $|a_{\vec{k},\sigma}|^2$, carrying energy $\hbar\omega|a_{\vec{k},\sigma}|^2$ and momentum $\hbar\vec{k}|a_{\vec{k},\sigma}|^2$. The appearance of $\hbar$ at this point does not indicate a truly quantum mechanical result, but rather our peculiar normalization — at this juncture — of the complex amplitudes.

5.8 Angular momentum, revisited

The angular momentum of the electromagnetic field,

$$\vec{L} = \frac{1}{4\pi c} \int_{\text{all space}} d^3x \, \left[\vec{x} \times (\vec{E} \times \vec{B}) \right] \tag{5.88}$$

is a little more complicated than the energy and momentum since it can be viewed as being composed of a reference frame dependent "orbital" part and an intrinsic "spin" part. The spin part has a relativistically invariant meaning and will correspond to the helicity of photons, so we want to have a look at it here.

To decompose the angular momentum into its orbital and spin parts we will need a slightly complicated vector identity:

$$[\vec{x} \times (\vec{E} \times \vec{B})]_i = (\vec{E} \times \vec{A})_i + \vec{E} \cdot (\vec{x} \times \nabla)_i \vec{A} + \nabla \cdot [(\vec{A} \times \vec{x})_i \vec{E}], \tag{5.89}$$

where the index i labels the Cartesian components of the indicated vectors. The only good way I know to derive this identity is to use $\vec{B} = \nabla \times \vec{A}$ and then use the fact that Cartesian components of the cross product and curl can be expressed in terms of the alternating symbol:

$$(\vec{C} \times \vec{D})_i = \epsilon_{ijk} C^j D^k, \quad (\nabla \times \vec{D})_i = \epsilon_i{}^j{}_k \frac{\partial}{\partial x^j} D^k, \tag{5.90}$$

where repeated indices are summed over. Then one has to use the identity

$$\epsilon^{ijk} \epsilon_{ilm} = \delta^j_l \delta^k_m - \delta^k_l \delta^j_m. \tag{5.91}$$

With this identity in hand, one can decompose the angular momentum integrand into the three terms shown in (5.89). The last term in (5.89) is a divergence, so when it appears in the integral (5.88) it can be transformed to the boundary of the region – spatial infinity – by the divergence

theorem. Assuming that $\vec{A}$ vanishes rapidly enough at infinity this term will not contribute. The second term in (5.89) depends upon a reference point and can be viewed as giving the density of "orbital" angular momentum. The first term in (5.89) does not depend upon a reference point and yields the "intrinsic" or "spin" angular momentum density.

Let us express the intrinsic part of the angular momentum,

$$\vec{J} = \frac{1}{4\pi c}\int_{\text{all space}} d^3x\,(\vec{E}\times\vec{A}), \tag{5.92}$$

in terms of the Fourier amplitudes. Substituting $\vec{E} = -\frac{1}{c}\frac{\partial \vec{A}}{\partial t}$ and then using the formula (5.78) we get

$$\vec{J} = i\hbar\int_{\mathbf{R}^3} d^3k\,\frac{\vec{k}}{k}\left[a_1^*(\vec{k})a_2(\vec{k}) - a_2^*(\vec{k})a_1(\vec{k})\right]. \tag{5.93}$$

To put this result in a more useful form, introduce the following complex basis for the vector space orthogonal to $\vec{k}$:

$$\vec{\varepsilon}_+(\vec{k}) = -\frac{1}{\sqrt{2}}(\vec{\varepsilon}_1(\vec{k}) + i\vec{\varepsilon}_2(\vec{k}) \tag{5.94}$$

$$\vec{\varepsilon}_-(\vec{k}) = \frac{1}{\sqrt{2}}(\vec{\varepsilon}_1(\vec{k}) - i\vec{\varepsilon}_2(\vec{k}) \tag{5.95}$$

and define corresponding amplitudes

$$a_+(\vec{k}) = -\frac{1}{\sqrt{2}}(a_1(\vec{k}) - ia_2(\vec{k})) \tag{5.96}$$

$$a_-(\vec{k}) = \frac{1}{\sqrt{2}}(a_1(\vec{k}) + ia_2(\vec{k})) \tag{5.97}$$

from which it follows that

$$a_1(\vec{k})\vec{\varepsilon}_1(\vec{k}) + a_2(\vec{k})\vec{\varepsilon}_2(\vec{k}) = a_+(\vec{k})\vec{\varepsilon}_+(\vec{k}) + a_-(\vec{k})\vec{\varepsilon}_-(\vec{k}), \tag{5.98}$$

and

$$i(a_1^*(\vec{k})a_2(\vec{k}) - a_2^*(\vec{k})a_1(\vec{k})) = a_+^*(\vec{k})a_+(\vec{k}) - a_-^*(\vec{k})a_-(\vec{k}). \tag{5.99}$$

The functions $a_\pm(\vec{k})$ are the complex amplitudes for plane waves with *circular polarization*. In terms of these amplitudes we have

$$\vec{J} = \hbar\int_{\mathbf{R}^3} d^3k\,\frac{\vec{k}}{k}(a_+^*(\vec{k})a_+(\vec{k}) - a_-^*(\vec{k})a_-(\vec{k})). \tag{5.100}$$

The clockwise and counter-clockwise amplitudes for each $\vec{k}$ contribute equally in magnitude but oppositely in sign to the "spin" part of the angular momentum. For each $\vec{k}$ the intrinsic angular momentum density vector is directed along the wave vector $\vec{k}$; its magnitude can be shown to be invariant under a change of inertial reference frame. This relativistic invariant (divided by $\hbar$) is known as the *helicity* of the plane wave with wave vector $\vec{k}$. I will denote the helicity of a wave with wave vector $\vec{k}$ by

$$\mathcal{J}(\vec{k}) = a_+^*(\vec{k})a_+(\vec{k}) - a_-^*(\vec{k})a_-(\vec{k}). \tag{5.101}$$

5.9 Periodic boundary conditions

As shown above, the states of the electromagnetic field (in the absence of sources) can be identified with a collection of harmonic oscillators, which are playing the role of normal modes of vibration for the system. This interpretation is somewhat figurative since the normal modes are (1) infinite in number and (2) labeled by a continuous set of variables $\vec{k}$. There is nothing to be done about (1); in contrast to a mechanical system, a field has an infinite number of degrees of freedom. But there is something we can do about (2) which will be convenient in the next chapter.

Imagine putting the electromagnetic field in a large but finite cube of edge length L

$$0 \le x \le L, \quad 0 \le y \le L, \quad 0 \le z \le L, \tag{5.102}$$

and imposing periodic boundary conditions. The idea is that for large enough L any physical results pertaining to a bounded region will differ insignificantly from the results obtained using (5.102).[10] We can use this artifice to render the normal modes into a discrete (if infinite) set. To be sure, it is not necessary to do this. I just thought it might make the upcoming construction of photons a little easier to assimilate if we kept the description a little closer to that of a mechanical system by using the artifice of periodic boundary conditions. This will also allow us to avoid a couple of technical digressions. So let me summarize here what happens to the normal modes in this case.

Working in a cube of edge length L and periodic boundary conditions the solutions to the source-free Maxwell equations are of the form

$$\vec{A}(\vec{x},t) = \frac{1}{L^{3/2}} \sum_{\vec{k}} \sum_{\sigma=1}^{2} \sqrt{\frac{2\pi\hbar c}{k}} [a_\sigma(\vec{k})\, \vec{\varepsilon}_\sigma(\vec{k})\, e^{i\vec{k}\cdot\vec{x} - i\omega(k)t} + a^*_\sigma(\vec{k})\, \vec{\varepsilon}_\sigma(\vec{k})\, e^{-i\vec{k}\cdot\vec{x} + i\omega(k)t}], \tag{5.103}$$

where the wave vectors are restricted to

$$k_x = \frac{2\pi n_x}{L}, \quad n_x = 0, \pm 1, \pm 2, \ldots, \tag{5.104}$$

$$k_y = \frac{2\pi n_y}{L}, \quad n_y = 0, \pm 1, \pm 2, \ldots, \tag{5.105}$$

$$k_z = \frac{2\pi n_z}{L}, \quad n_z = 0, \pm 1, \pm 2, \ldots, \tag{5.106}$$

and

$$\vec{k} \neq 0. \tag{5.107}$$

The solution is similar to (5.80), but with integrals over Fourier space changed to summations over the three components of $\vec{k}$:

$$\frac{1}{(2\pi)^{3/2}} \int_{\mathbf{R}^3} d^3k \quad \longrightarrow \quad \frac{1}{L^{3/2}} \sum_{\vec{k}} \equiv \frac{1}{L^{3/2}} \sum_{n_x, n_y, n_z = -\infty}^{\infty}. \tag{5.108}$$

Of course, integrals over all space become integrals over the cube:

$$\int_{\text{all space}} d^3x = \int_0^L dx \int_0^L dy \int_0^L dz \tag{5.109}$$

[10] Other boundary conditions, *e.g.,* Dirichlet conditions, would serve as well. Periodic boundary conditions are advantageous since they allow for mode functions which are simple exponentials.

Notice that I have excluded the zero frequency mode via (5.107). The zero frequency mode represents a constant vector potential. A constant vector potential contributes nothing to the electromagnetic field since the electromagnetic field is constructed from derivatives of $\vec{A}$. Notice also that a gauge transformation by a function Λ in this finite volume setting will preserve the radiation gauge if and only if $\Lambda = const.$, which is a trivial transformation in that it does not change the vector potential. Consequently, with periodic boundary conditions and the exclusion of the zero frequency mode we obtain a unique representation of the electromagnetic field in terms of the radiation gauge potentials in this finite-volume setting.

5.10 Motion of a charged particle in an electromagnetic field

For a given electromagnetic field described by potentials $(\phi, \vec{A})$ the Lagrangian and Hamiltonian describing the motion of a non-relativistic particle with mass m and charge q is given by

$$L(\vec{x}, \vec{v}, t) = \frac{1}{2}mv^2 + \frac{q}{c}\vec{A}(\vec{x}, t) \cdot \vec{v} - q\phi(\vec{x}, t), \quad \vec{v} \equiv \frac{d\vec{x}(t)}{dt}, \tag{5.110}$$

$$H(\vec{x}, \vec{p}, t) = \frac{1}{2m}\left[\vec{p} - \frac{q}{c}\vec{A}(\vec{x}, t)\right]^2 + q\phi(\vec{x}, t). \tag{5.111}$$

The Euler-Lagrange equations yield the Lorentz force law:

$$m\frac{d\vec{v}}{dt} = q\vec{E}(\vec{x}, t) + \frac{q}{c}\vec{v} \times \vec{B}(\vec{x}, t). \tag{5.112}$$

The Hamilton equations,

$$\frac{d\vec{x}}{dt} = \frac{\partial H}{\partial \vec{p}} = \frac{1}{m}(\vec{p} - \frac{q}{c}\vec{A}), \tag{5.113}$$

$$\frac{d\vec{p}}{dt} = -\frac{\partial H}{\partial \vec{x}}. \tag{5.114}$$

are equivalent to (5.112).

The functional forms of both the Lagrangian and the Hamiltonian change when a gauge transformation is made. Despite this, the equations of motion evidently manage to be invariant under gauge transformations. This can be understood in the Lagrangian formalism by the fact that under a gauge transformation the Lagrangian changes only by addition of a total time derivative, which does not alter the Euler-Lagrange equations. From the Hamiltonian perspective, the change in the Hamiltonian corresponds to that induced by a canonical transformation, and canonical transformations yield equivalent equations of motion.

5.11 Problems

1. Show that the magnetic field satisfies the same wave equation as the electric field.

2. The *constraint equations* are the two scalar Maxwell equations (5.2), (5.3), which do not involve time derivatives. The *evolution equations* are the remaining Maxwell equations (5.4), (5.5), involving the time derivatives. Show that if the constraint equations are satisfied at any one time, and the electromagnetic field at any later time is obtained by solving the

evolution equations with the sources satisfying the continuity equation, then the constraint equations are automatically satisfied at any later time. Thus the constraints can be viewed only as restrictions on the initial data for the evolution equations. (*Hint:* Show that the time derivatives of $\nabla \cdot \vec{E} - 4\pi\rho$ and $\nabla \cdot \vec{B}$ vanish.)

3. Verify that the momentum and angular momentum are conserved when the source-free Maxwell equations are satisfied.

4. A quantum particle with mass m and charge q is moving in a given electromagnetic field represented by potentials $(\phi(\vec{x},t), \vec{A}(\vec{x},t))$. The motion of the particle is described by the Hamiltonian operator:
$$\hat{H} = \frac{1}{2m}\left[\hat{\vec{p}} - \frac{q}{c}\vec{A}(\hat{\vec{x}},t)\right]^2 + q\phi(\hat{\vec{x}},t). \tag{5.115}$$
Solve the eigenvalue problem for this Hamiltonian in the case of a static, uniform magnetic field $\vec{B} = B\hat{z}$, $B = \text{const.}$, which can be described by the potentials
$$\phi = 0, \quad \vec{A} = \frac{1}{2}\vec{B} \times \vec{x}. \tag{5.116}$$

5. Consider an infinite cylindrical region of radius R. Let $\vec{B}$ be a magnetic field which vanishes outside the cylinder; inside the cylinder $\vec{B}$ is uniform and is directed along the axis of the cylinder. (This is the magnetic field created by a solenoid.) Show that the vector potential cannot vanish everywhere outside the cylinder, even though $\vec{B} = 0$ there. Find a set of potentials to describe this magnetic field.

6. This problem makes use of the previous two problems. Let a particle of charge q and mass m move in a cylindrical region with its cylindrical coordinate restricted by $R_1 < r < R_2$. In the cylindrical region $r < R_1$ there is a static, uniform magnetic field directed along the axis of the cylinder. The magnetic field vanishes for $r > R_1$. Consider the eigenvalue equation for energies and stationary states. Show that the energies depend upon the magnetic field even though the particle has vanishing probability to be found in the region where the magnetic field is non-vanishing. (This is the *Aharanov-Bohm effect*. For you mathematical types, this effect shows the utility of *holonomy of a connection* in quantum mechanics.)

7. Show that if the electromagnetic field is expressed in terms of potentials as in (5.51) then the homogeneous Maxwell equations (5.3) and (5.5) are satisfied.

8. The *Coulomb field* is the electric field outside a static, spherically symmetric charge distribution. It is given by
$$\vec{E} = \frac{k}{r^3}\vec{x}, \quad \vec{B} = 0, \tag{5.117}$$
where k is a constant, $\vec{x}$ is the position relative to the center of symmetry and $r^2 = \vec{x}\cdot\vec{x}$. Find the vector potential in the radiation gauge which represents the Coulomb field. Show that it satisfies the wave equation (5.64).

9. Verify that (5.76) follows from (5.64).

10. Verify (5.87).

11. The quantity $\mathcal{L} = E^2 - B^2$ is a relativistic invariant. Express

$$L = \frac{1}{8\pi} \int_{\text{all space}} d^3x \, \mathcal{L}$$

in terms of the complex amplitudes defining the solution (5.78), and interpret it in terms of harmonic oscillator Lagrangians. In this way L serves as a Lagrangian for the electromagnetic field.

12. Derive the vector identity (5.89).

13. Derive the formula (5.100) for the helicity.

14. Consider an electromagnetic wave with a given wave vector $\vec{k}$. Show that the helicity (5.101) vanishes for a plane polarized wave, and that it is non-vanishing for a circularly polarized wave.

15. Consider an interval $[0, L]$ and the set of smooth, periodic functions on it. Show that when "integrated" against such functions, the following expression defines a delta function:

$$\delta(x, y) = \frac{1}{L} \sum_{n=0}^{\infty} e^{in(x-y)\frac{2\pi}{L}}. \tag{5.118}$$

16. In the context of periodic boundary conditions, show that the radiation gauge vector potential is uniquely determined by the electromagnetic field. (*Hint:* Show that a gauge transformation $\vec{A} \to \vec{A} + \nabla\Lambda$ which preserves the divergence condition $\nabla \cdot \vec{A} = 0$ is necessarily trivial.)

Chapter 6

What is a Photon?

We have now assembled all the ingredients to build a quantum theory of electromagnetism and, in particular, its manifestation in terms of photons. The way to proceed is clear: we have isolated the normal modes of the (source-free) electromagnetic field, all that remains to be done is to view them as quantum oscillators and see what happens. We shall see that the physical observables of the quantum normal modes provide them with the usual particle-like features we empirically associate with the phenomena of photons.

6.1 Hilbert space for the electromagnetic field

We have seen that associated to each wave vector $\vec{k}$ and polarization state σ there is an electromagnetic normal mode of oscillation with frequency $\omega(k) = kc$ and complex amplitude $a_\sigma(\vec{k})$. In the absence of sources, the behavior of the electromagnetic field is completely described by these normal modes. We will assume the normal modes are defined using periodic boundary conditions as in §5.9. We make the transition to quantum field theory by assuming these normal modes are quantum oscillators. We use the tensor product Hilbert space to describe states of all these oscillators. On this space, for each $\vec{k}$ and σ there will be ladder operators $\hat{a}_\sigma(\vec{k})$, $\hat{a}^\dagger_\sigma(\vec{k})$. All these ladder operators are just multiple copies of the usual ladder operators for a single oscillator. In particular we have the commutation relations

$$[\hat{a}_\sigma(\vec{k}), \hat{a}_\tau(\vec{l})] = 0 = [\hat{a}^\dagger_\sigma(\vec{k}), \hat{a}^\dagger_\tau(\vec{l})], \quad [\hat{a}_\sigma(\vec{k}), \hat{a}^\dagger_\tau(\vec{l})] = \delta_{k_x,l_x}\delta_{k_y,l_y}\delta_{k_z,l_z}\delta_{\sigma\tau}\hat{1} \equiv \delta_{\vec{k},\vec{l}}\delta_{\sigma\tau}\hat{1}. \tag{6.1}$$

For each $\vec{k}$ and σ there is a Hilbert space of square integrable functions and a Hamiltonian $\hat{H}_\sigma(\vec{k}) = \hbar\omega(k)\hat{a}^\dagger_\sigma(\vec{k})\hat{a}_\sigma(\vec{k})$. For each $\vec{k}$ and σ there is a basis of energy eigenvectors; the corresponding ladder operators act on this basis by adding and subtracting quanta of energy $\hbar\omega(k)$. We can use this basis to induce a basis for the tensor product over all the oscillator Hilbert spaces, which is the state space for the composite system of normal modes, *i.e.,* the electromagnetic field. The elements of this basis can be labeled by assigning a non-negative integer $n_\sigma(\vec{k})$ to each pair $(\vec{k}, \sigma)$. Given a function $\mathbf{n}$ mapping $(\vec{k}, \sigma)$ to the integers we could denote the corresponding basis vector by $|\mathbf{n}\rangle$. This is a state where (with statistical certainty) the normal mode labeled by $\vec{k}$, σ has energy $n_\sigma(\vec{k})\hbar kc$. This energy basis is orthonormal. As usual, the state of the system can be expressed as a

normalized linear combination of vectors from the energy basis. All this is just a slightly elaborate version of the quantum theory of multiple independent oscillators.

6.2 The Hamiltonian. Stationary states

In accord with (3.24), the total energy, or the Hamiltonian, of the system is represented by the operator

$$\hat{H} = \sum_{\vec{k},\sigma} \hbar kc \; \hat{a}^\dagger_\sigma(\vec{k})\hat{a}_\sigma(\vec{k}), \tag{6.2}$$

where the various tensor products with the identity operator are suppressed. This is the quantum version of (5.86). The ground state of the system has all normal modes in their ground state. This is the state defined by the basis element with $n_\sigma(\vec{k}) = 0$, $\forall \vec{k}$, $\sigma = 1, 2$ and we denote it by $|0\rangle$. This is *not* the zero vector! Rather it is the product of oscillator ground states (with Gaussian wave functions), one ground state for each normal mode. The ground state of the electromagnetic field is mapped to zero by all the lowering operators:

$$a_\sigma(\vec{k})|0\rangle = 0, \quad \forall \; \vec{k}, \; \sigma = 1, 2. \tag{6.3}$$

Consequently, the ground state is an eigenvector of $\hat{H}$ with eigenvalue zero,

$$\hat{H}|0\rangle = 0, \tag{6.4}$$

representing a state of the electromagnetic field where the energy vanishes with statistical certainty. Classically, this lowest energy state corresponds to the vector potential $\vec{A} = 0$. Quantum mechanically, things are more interesting, as we shall see.

The first excited state of the quantum electromagnetic field has one quantum of energy, $2\pi\hbar c/L$, in one of the 6 quantum oscillators having $k = \frac{2\pi}{L}$. A first excited state vector can be obtained by

$$|1_{\vec{k},\sigma}\rangle = a^\dagger_\sigma(\vec{k})|0\rangle, \tag{6.5}$$

where $k = \frac{2\pi}{L}$. You can easily check that this is an eigenvector of $\hat{H}$ with eigenvalue $2\pi\hbar c/L$. This eigenvalue has a six-fold degeneracy. The state $\hat{a}^\dagger_\sigma(\vec{k})|0\rangle$ has one quantum of energy $\hbar\omega(k)$ in the quantum mode with wave vector $\vec{k}$ and polarization σ.

One can continue adding quanta to normal modes via application of $\hat{a}^\dagger_\sigma(\vec{k})$, with various choices for $\vec{k}$ and σ, and build all the stationary states of the system. For example, states with a single quantum in each of the distinct normal modes defined by $(\sigma_1, \vec{k}_1)$, $(\sigma_2, \vec{k}_2)$, $\ldots$, $(\sigma_l, \vec{k}_l)$ are of the form

$$|1_{\vec{k}_1,\sigma_1}, 1_{\vec{k}_2,\sigma_2}, \ldots 1_{\vec{k}_l,\sigma_l}\rangle = a^\dagger_{\sigma_1}(\vec{k}_1)a^\dagger_{\sigma_2}(\vec{k}_2)\cdots a^\dagger_{\sigma_l}(\vec{k}_l)|0\rangle, \tag{6.6}$$

and satisfy the eigenvalue equation

$$\hat{H}|1_{\vec{k}_1,\sigma_1}, 1_{\vec{k}_2,\sigma_2}, \ldots 1_{\vec{k}_l,\sigma_l}\rangle = \hbar\left[\omega(k_1) + \omega(k_2) + \cdots + \omega(k_l)\right]|1_{\vec{k}_1,\sigma_1}, 1_{\vec{k}_2,\sigma_2}, \ldots 1_{\vec{k}_l,\sigma_l}\rangle. \tag{6.7}$$

Here is a state vector representing n quanta in the normal mode labeled by $(\vec{k}, \sigma)$:

$$|n_{\sigma,\vec{k}}\rangle = \frac{1}{\sqrt{n!}}\left(\hat{a}_\sigma(\vec{k})^\dagger\right)^n|0\rangle \tag{6.8}$$

This is an eigenvector of the Hamiltonian with eigenvalue $n\hbar kc$.

The Hamiltonian (6.2) is self-adjoint and has a basis of eigenvectors such as described in the previous paragraph. Every state of the quantum electromagnetic field can be expressed as a superposition of these eigenvectors. As always in quantum mechanics, energy eigenvectors correspond to stationary state solutions of the Schrödinger equation, which do not have any observable time dependent behavior. Time dependent behavior occurs in superpositions of stationary states with different energies.

6.3 Momentum and Helicity

Motivated by (5.86) and (5.87), let us now define a triplet of operators representing linear momentum of the quantum electromagnetic field by

$$\hat{\vec{P}} = \sum_{\vec{k},\sigma} \hbar\vec{k}\, \hat{a}^\dagger_\sigma(\vec{k})\hat{a}_\sigma(\vec{k}). \tag{6.9}$$

The ground state $|0\rangle$ described in the previous section is easily seen to be an eigenstate of these operators with eigenvalue zero:

$$\hat{\vec{P}}|0\rangle = 0. \tag{6.10}$$

The excited states $|1_{\vec{k},\sigma}\rangle$ are eigenvectors of $\hat{\vec{P}}$ with eigenvalue $\hbar\vec{k}$,To verify this, you can calculate as follows:

$$\hat{\vec{P}}|1_{\vec{k},\sigma}\rangle = \left\{\sum_{\vec{l}}\sum_{\tau} \hbar\vec{l}\,\hat{a}^\dagger_\tau(\vec{l})\hat{a}_\tau(\vec{l})\right\} a^\dagger_\sigma(\vec{k})|0\rangle \tag{6.11}$$

$$= \sum_{\vec{l}}\sum_{\tau} \hbar\vec{l}\hat{a}^\dagger_\tau(\vec{l})[\hat{a}_\tau(\vec{l}), a^\dagger_\sigma(\vec{k})]|0\rangle \tag{6.12}$$

$$= \sum_{\vec{l}}\sum_{\tau} \hbar\vec{l}\hat{a}^\dagger_\tau(\vec{l})\delta_{\vec{l},\vec{k}}\delta_{\sigma,\tau}|0\rangle \tag{6.13}$$

$$= \hbar\vec{k}a^\dagger_\sigma(\vec{k})|0\rangle \tag{6.14}$$

$$= \hbar\vec{k}|1_{\vec{k},\sigma}\rangle. \tag{6.15}$$

More generally, in states such as (6.6), (6.8) with various normal modes excited we get

$$\hat{\vec{P}}|1_{\vec{k}_1,\sigma_1}, 1_{\vec{k}_2,\sigma_2}, \ldots 1_{\vec{k}_l,\sigma_l}\rangle = \hbar\left(\vec{k}_1 + \vec{k}_2 + \cdots + \vec{k}_l\right)|1_{\vec{k}_1,\sigma_1}, 1_{\vec{k}_2,\sigma_2}, \ldots 1_{\vec{k}_l,\sigma_l}\rangle, \tag{6.16}$$

$$\hat{\vec{P}}|n_{\sigma,\vec{k}}\rangle = n\hbar\vec{k}|n_{\sigma,\vec{k}}\rangle. \tag{6.17}$$

A straightforward calculation shows that $[\hat{\vec{P}}, \hat{H}] = 0$, which means there exists a simultaneous basis of eigenvectors for energy and momentum. This basis is the set of stationary states just described above and in the previous section.

Similarly, motivated by (5.100) and (5.101), let us define the operator representing intrinsic angular momentum,

$$\hat{J} = \hbar \sum_{\vec{k}} \frac{\vec{k}}{k} \left[\hat{a}_+^\dagger(\vec{k}) \hat{a}_+(\vec{k}) - \hat{a}_-^\dagger(\vec{k}) \hat{a}_-(\vec{k}) \right], \tag{6.18}$$

and helicity of a normal mode labeled by $\vec{k}$:

$$\hat{\mathcal{J}}_{\vec{k}} = \hat{a}_+^\dagger(\vec{k}) \hat{a}_+(\vec{k}) - \hat{a}_-^\dagger(\vec{k}) \hat{a}_-(\vec{k}), \tag{6.19}$$

where (*cf* (5.97))

$$\hat{a}_+(\vec{k}) = -\frac{1}{\sqrt{2}} (\hat{a}_1(\vec{k}) - i\hat{a}_2(\vec{k})) \tag{6.20}$$

$$\hat{a}_-(\vec{k}) = \frac{1}{\sqrt{2}} (\hat{a}_1(\vec{k}) + i\hat{a}_2(\vec{k})). \tag{6.21}$$

You can check that the intrinsic angular momentum and the helicity operators commute with the Hamiltonian and the momentum, so all stationary states can be chosen to be states of definite momentum and definite helicity. Aside from the ground state, which has no intrinsic angular momentum with certainty, the stationary states with quanta in normal modes associated to $\vec{\epsilon}_1$, $\vec{\epsilon}_2$ are, however, *not* eigenvectors of the helicity. For example,

$$\hat{\mathcal{J}}\, \hat{a}_1^\dagger(\vec{k}) |0\rangle = \sqrt{2} \hat{a}_2^\dagger(\vec{k}) \, |0\rangle. \tag{6.22}$$

Linear combinations of stationary states (6.6) which correspond to exciting modes with a fixed $\vec{k}$ and with circular polarization $\vec{\epsilon}_\pm$ *are* helicity eigenvectors and are still eigenvectors of energy and momentum. These states can be obtained by applying $\hat{a}_\pm(\vec{k})$ to the ground state. It is straightforward to compute

$$[\hat{\mathcal{J}}_{\vec{k}}, \hat{a}_\pm^\dagger(\vec{k})] = \pm \hat{a}_\pm^\dagger(\vec{k}), \tag{6.23}$$

showing that adding quanta into the normal modes associated with $\vec{\epsilon}_\pm$ changes the helicity by ± 1 with statistical certainty. In particular, the states with one quantum in the normal mode associated to $\vec{\epsilon}_\pm$ has helicity ± 1 with certainty:

$$\hat{\mathcal{J}}_{\vec{k}} \hat{a}_\pm^\dagger(\vec{k}) |0\rangle = \pm \hat{a}_\pm^\dagger(\vec{k}) |0\rangle. \tag{6.24}$$

The operators $\hat{a}_\pm(\vec{k})$ and $\hat{a}_\pm^\dagger(\vec{k})$ satisfy the algebra of creation and annihilation operators and can be used to build a basis of energy and momentum eigenvectors by repeatedly applying all possible creation operators $a_\pm^\dagger(\vec{k})$ to the ground state. In this way one can build a basis of energy, momentum, and helicity eigenvectors. Of course, as usual in quantum mechanics, a general state of the system is a superposition over this basis and so does not in general have a statistically sharp value for energy, or momentum, or helicity.

6.4 Interpretation in terms of photons

I pointed out in §4.6 that one can identify the Hilbert space of an infinite collection of harmonic oscillators with a Fock space of bosons. It is this identification which provides the interpretation of

the quantum electromagnetic field in terms of photons. Recall that the Fock space interpretation has the number of quanta in a given oscillator energy eigenstate corresponding to the number of particles in a corresponding 1-particle state. So, to introduce the photon interpretation we need a correspondence between oscillator states and 1-particle states. This is done as follows.

First, identify the ground state of the oscillator system $|0\rangle$ with the vacuum state in the Fock space. Next, consider a state with one quantum of energy in the normal mode defined by $\vec{k}$ and with circular polarization. Mathematically it can be defined by

$$|1_{\vec{k},\pm}\rangle = a^\dagger_\pm(\vec{k})|0\rangle. \tag{6.25}$$

This vector is identified with a 1-particle state in Fock space characterized by statistically sharp values of energy ($\hbar kc$), momentum ($\hbar\vec{k}$), and helicity (± 1). Of course all these observables are in accord with the standard phenomenology of photons. The 1-particle Hilbert space is spanned by the states (6.25) as $\vec{k}$ and $\sigma = \pm$ range over all possible values. This means the Hilbert space of states of a single photon can be identified with the vector space of pairs of functions:[1]

$$|\psi\rangle_1 = \begin{pmatrix} \psi_+(\vec{k}) \\ \psi_-(\vec{k}) \end{pmatrix}, \tag{6.26}$$

with the scalar product

$${}_1\langle\psi|\phi\rangle_1 = \sum_{\vec{k}} \begin{pmatrix} \psi^*_+(\vec{k}) & \psi^*_-(\vec{k}) \end{pmatrix} \begin{pmatrix} \phi_+(\vec{k}) \\ \phi_-(\vec{k}) \end{pmatrix}. \tag{6.27}$$

This specification of the 1-particle Hilbert space determines the (symmetric) Fock space, as discussed in Chapter 4.

The identification of Fock space with the Hilbert space of quantum normal modes has an occupation number state for n-photons with energy, ($\hbar kc$), momentum ($\hbar\vec{k}$), and helicity (± 1) being represented by the state vector

$$|n(\vec{k},\pm)\rangle = \frac{1}{\sqrt{n!}}\hat{a}^\dagger_\pm(\vec{k})\hat{a}^\dagger_\pm(\vec{k})\cdots\hat{a}^\dagger_\pm(\vec{k})|0\rangle, \tag{6.28}$$

where there are n creation operators. In this correspondence, adding a quantum of energy to the normal mode specified by wave vector $\vec{k}$ and helicity ± 1 is identified with adding a photon to the system in the state characterized by energy $\hbar kc$, momentum $\hbar\vec{k}$, and helicity ± 1. In this regard, notice that the energy, momentum and helicity operators (6.2), (6.9), (6.19) are precisely in the form of additive one particle observables as discussed in §6.9.

It is worth noting that it is possible to have photons which are *not* in states of definite energy, momentum, or helicity. This is done by taking linear combinations over the states studied above. For example, consider the state vector given by

$$|f\rangle = \sum_{\vec{k}} \left(f_+(\vec{k})|1_{\vec{k},+}\rangle + f_-(\vec{k})|1_{\vec{k},-}\rangle \right), \tag{6.29}$$

where $f_\pm(\vec{k})$, $\sigma = 1, 2$, is a pair of functions satisfying

$$\sum_{\vec{k}} \left(|f_+(\vec{k})|^2 + |f_+(\vec{k})|^2 \right) = 1. \tag{6.30}$$

[1]The subscript "1" is just to remind us that we are working with the 1-particle Hilbert space, not the whole Fock space.

This state is an eigenvector of the total number operator

$$\hat{N} = \sum_{\vec{k}} \sum_{\sigma} \hat{a}^{\dagger}_{\sigma}(\vec{k}) \hat{a}_{\sigma}(\vec{k}) \tag{6.31}$$

with eigenvalue 1, so it represents a single photon with certainty. This photon, however, has a statistical distribution of momenta, energies, helicities since

$$|\langle 1_{\vec{k},\pm}|f\rangle|^2 = |f_{\pm}(\vec{k})|^2 \tag{6.32}$$

is the probability that the photon has momentum $\hbar\vec{k}$, energy $\hbar kc$ and helicity ± 1.

It is also possible to have states where even the number of photons is uncertain! States with a statistically definite number of photons will be eigenstates of the total number operator (6.31). States with uncertainty in the number of photons can be obtained by superposing states with different numbers of photons. We shall see a little later that the coherent states which present classical behavior (in macroscopic circumstances) for the electromagnetic field are states with a statistically uncertain number of photons.

6.5 Time evolution

In our discussion thus far I have been tacitly working in the Schrödinger picture of dynamics. It was not necessary to think much about this since we have been focusing on stationary states and on observables like energy, momentum, and helicity, which are conserved quantities, so time evolution has not been in play. Let's briefly think about dynamics a bit more explicitly.

Since the quantum electromagnetic field was constructed via the quantum oscillator, the time evolution of the field follows the patterns established there. In the Heisenberg picture, the operators $\hat{a}_{\sigma}(\vec{k}, t)$ – representing ladder operators for normal modes and, equivalently, creation and annihilation operators for photons – evolve according to the Heisenberg equations:

$$i\hbar \frac{d}{dt} \hat{a}_{\sigma}(\vec{k}, t) = [\hat{a}_{\sigma}(\vec{k}, t), \hat{H}] = \hbar\omega(k) a_{\sigma}(\vec{k}, t), \tag{6.33}$$

so that

$$\hat{a}_{\sigma}(\vec{k}, t) = e^{-i\omega(k)t} \hat{a}_{\sigma}(\vec{k}). \tag{6.34}$$

This is in compete accord with the solution to the Maxwell equations (5.78), whence we recover the usual quantum mechanical result that the classical equations of motion determine the Heisenberg picture observables (at least for linear equations of motion). The corresponding Schrödinger picture operators can be obtained by evaluating the Heisenberg picture operators at $t = 0$.

6.6 The electromagnetic field operator

Using the relation between classical and quantum observables for a harmonic oscillator, and keeping in mind the classical-quantum relation obtained via coherent states, we define the operator representing the vector potential (in the radiation gauge) to be

$$\hat{\vec{A}}(\vec{x}, t) = \frac{1}{L^{3/2}} \sum_{\vec{k}} \sum_{\sigma=1}^{2} \sqrt{\frac{2\pi\hbar c}{k}} [\hat{a}_{\sigma}(\vec{k})\, \vec{\varepsilon}_{\sigma}(\vec{k})\, e^{i\vec{k}\cdot\vec{x} - i\omega(k)t} + \hat{a}^{\dagger}_{\sigma}(\vec{k})\, \vec{\varepsilon}_{\sigma}(\vec{k})\, e^{-i\vec{k}\cdot\vec{x} + i\omega(k)t}]. \tag{6.35}$$

Similarly we define the operators representing the electric and magnetic fields to be

$$\hat{\vec{E}}(\vec{x},t) = -\frac{1}{c}\frac{\partial}{\partial t}\hat{\vec{A}}(\vec{x},t), \quad \hat{\vec{B}}(\vec{x},t) = \nabla \times \hat{\vec{A}}(\vec{x},t). \tag{6.36}$$

These formulas define the Heisenberg picture operators representing the field observables. As is fitting for Heisenberg picture operators, $\hat{\vec{A}}(\vec{x},t)$, $\hat{\vec{E}}(\vec{x},t)$, and $\hat{\vec{B}}(\vec{x},t)$ satisfy the relevant equations of motion, *e.g.,* $\hat{\vec{A}}(\vec{x},t)$ satisfies the source-free Maxwell equations in the radiation gauge (5.64). The Schrödinger picture version of these operators can be obtained by setting $t = 0$. We shall see in §7 that this definition of the electromagnetic field operators will serve nicely to explain phenomena like spontaneous emission.

It is a significant feature of the quantum electromagnetic field that the observable representing the number of photons is not compatible with the electromagnetic field itself. Mathematically, the operators representing $\vec{E}$ and $\vec{B}$ do not commute with the number operators,

$$\hat{N}_{\vec{k},\sigma} = \hat{a}^\dagger_\sigma(\vec{k})\hat{a}_\sigma(\vec{k}), \tag{6.37}$$

for photons of a given type. It is not hard to calculate

$$[\hat{\vec{A}}(\vec{x},t), \hat{N}_{\vec{k},\sigma}] = \frac{1}{L^{3/2}}\sqrt{\frac{2\pi\hbar c}{k}}[\hat{a}_\sigma(\vec{k})\,\vec{\varepsilon}_\sigma(\vec{k})\,e^{i\vec{k}\cdot\vec{x}-i\omega(k)t} - \hat{a}^\dagger_\sigma(\vec{k})\,\vec{\varepsilon}_\sigma(\vec{k})\,e^{-i\vec{k}\cdot\vec{x}+i\omega(k)t}]. \tag{6.38}$$

By differentiating with respect to space and time one can compute commutators of the number operator with the electromagnetic field and check that it is non-vanishing. For example,

$$[\hat{\vec{E}}(\vec{x},t), \hat{N}_{\vec{k},\sigma}] = \frac{i}{L^{3/2}}\sqrt{2\pi\hbar\omega(k)}[\hat{a}_\sigma(\vec{k})\,\vec{\varepsilon}_\sigma(\vec{k})\,e^{i\vec{k}\cdot\vec{x}-i\omega(k)t} + \hat{a}^\dagger_\sigma(\vec{k})\,\vec{\varepsilon}_\sigma(\vec{k})\,e^{-i\vec{k}\cdot\vec{x}+i\omega(k)t}]. \tag{6.39}$$

This is actually not so surprising. The Fourier modes of the electric field and the vector potential are the momentum and position observables for a quantum oscillator[2], and these have analogous commutation relations with the number operator. But the implication *is* somewhat striking: even in the state with no photons – the vacuum state – there will be a non-zero a probability for finding an electromagnetic field! You can see this by letting $\hat{\vec{E}}$ act on the vacuum and checking that the vacuum is not an eigenstate of $\hat{\vec{E}}$. As we shall explore a little in §6.8, in settings where the electromagnetic field has a relatively sharp "quasi-classical" value, statistically speaking, the quantum field will be in a state representing a superposition of states corresponding to many different numbers of photons. (Even for a single oscillator a state with, say, a well localized displacement is an infinite superposition of energy eigenstates.) Thus the classical electromagnetic field – even something as simple as a classical plane wave – is not simply described in terms of a single photon. This is, of course, why it took so long for the particle features of the field to be discovered.[3]

From the Fock space point of view, we see that the vector potential (6.35) is built from creation and annihilation operators $(\hat{a}_\pm(\vec{k}), \hat{a}^\dagger_\pm(\vec{k}))$ for particles in momentum eigenstates (and helicity eigenstates) at time t. It is tempting to then interpret $\hat{\vec{A}}(\vec{x},t)$ in terms of field operators, creating and annihilating photons at the location $\vec{x}$ at time t. As I will discuss in a little more detail later, while the momentum space interpretation of photons is straightforward, defining an unambiguous notion of position of a photon is problematic.

[2]This analogy can be seen already at the level of classical field theory, see *e.g.,* reference [5].

[3]By the same token, since the classical limit of the Schrödinger equation involves one or more particles, one can see why it took so long for the wave features of matter to be discovered.

6.7 Vacuum fluctuations

Let us follow up on the observation that the vacuum state has statistical fluctuations in its electromagnetic field. It is not hard to see how this comes about when you think in terms of quantum normal modes of vibration. The quantum fluctuations of the electromagnetic field in its vacuum state are just a manifestation of the familiar fact that the displacement and momentum for the oscillators making up the normal modes have statistical uncertainties in the oscillator ground state. To see how this works in detail from the quantum field point of view, we will consider the vacuum expectation value of an average of the electric field over a given volume at some fixed time. We shall define the corresponding operator by

$$\hat{\vec{E}}[f] = \int_{\text{all space}} d^3x\, f(\vec{x})\hat{\vec{E}}(\vec{x}), \tag{6.40}$$

where $\hat{\vec{E}} = -\frac{1}{c}\frac{\partial \hat{\vec{A}}}{\partial t}$, with $\hat{\vec{A}}$ given in (6.35), and $f(\vec{x})$ is any function we like. By choosing different functions we can study the values or the average values of $\vec{E}$ in any region we like. Mathematically, this averaging procedure will help us control singular quantities which can appear when considering quantum field values at a single point. Physically, when we measure the electric field we are at best measuring an average over some region and the operator (6.40) is a good mathematical model for that average observable.

Let us investigate the statistical properties of the averaged electric field in the vacuum state. Because this state is a stationary state, all such properties will be time independent and we can perform the calculations equally well in the Schrödinger or Heisenberg pictures. The vacuum expectation value of the average electric field vanishes,

$$\langle 0|\hat{\vec{E}}[f]|0\rangle = 0, \tag{6.41}$$

because it is a superposition of terms like

$$\langle 0|a_{\vec{k},\sigma}|0\rangle = 0 = \langle 0|a^\dagger_{\vec{k},\sigma}|0\rangle. \tag{6.42}$$

The vacuum expectation value of the square of this operator will then determine the statistical uncertainty of the electric field in the vacuum state. We have

$$\langle 0|\hat{\vec{E}}[f]\cdot\hat{\vec{E}}[g]|0\rangle = \int_{\text{all space}} d^3x\, d^3y f(\vec{x})g(\vec{y})\frac{1}{L^3}\sum_{\vec{k}} 4\pi\,\hbar kc\, e^{i\vec{k}\cdot(\vec{x}-\vec{y})}. \tag{6.43}$$

Evidently, the spatial integrals compute the Fourier transforms of the averaging functions. I want to show you that, for an average over a small enough volume $\sim \lambda^3$, $\lambda << L$, this expectation value is a pure number times $\frac{\hbar c}{\lambda^4}$.

To have an analytically tractable computation, let us suppose we average with a normalized Gaussian function:

$$f(\vec{x}) = g(\vec{x}) = \frac{1}{(\lambda\sqrt{2\pi})^3} e^{-\frac{1}{2\lambda^2}(x^2+y^2+z^2)}, \tag{6.44}$$

where λ sets the length scale of the volume about the origin where the electric field is appreciably sampled. The Fourier transform of a Gaussian is a Gaussian. For large enough L (compared to λ) we have

$$\int_{\text{all space}} d^3x\, f(\vec{x})e^{i\vec{k}\cdot\vec{x}} \approx e^{-\frac{1}{2}\lambda^2k^2} \tag{6.45}$$

This yields

$$\langle 0|\hat{E}[f]^2|0\rangle \equiv \langle 0|\hat{\vec{E}}[f]\cdot\hat{\vec{E}}[f]|0\rangle \approx \frac{1}{L^3}\sum_{\vec{k}} 4\pi\hbar kc\, e^{-\lambda^2 k^2}. \tag{6.46}$$

For small enough λ (compared to L) this sum is dominated by $k >> 1$, where we can approximate the sum by an integral:

$$\frac{1}{L^3}\sum_{\vec{k}} 4\pi\hbar kc\, e^{-\lambda^2 k^2} \approx \left(\frac{1}{2\pi}\right)^3 \int_{\mathbf{R}^3} d^3k\, 4\pi\hbar kc\, e^{-\lambda^2 k^2}. \tag{6.47}$$

With

$$\int_{\mathbf{R}^3} d^3k\, k e^{-\lambda^2 k^2} = \frac{2\pi}{\lambda^4}, \tag{6.48}$$

we get

$$\langle 0|(\hat{E}[f])^2|0\rangle \approx \frac{1}{\pi}\frac{\hbar c}{\lambda^4}. \tag{6.49}$$

We see that in the vacuum state the statistical uncertainty in the electric field averaged over a small volume is non-zero and increases without bound as the volume decreases. An analogous calculation can be made for the magnetic field with the same result. As already mentioned, this phenomenon is analogous to the "zero point fluctuations" of the position and momentum of an oscillator in its ground state.[4] One says that "vacuum fluctuations" in the electromagnetic field become larger and larger at shorter and shorter distance scales. In SI units the variance in the electromagnetic field from the above calculation is on the order of $10^{-26}/\lambda^4\ J/m^3$. For macroscopic situations these fluctuations are negligible. But, as you can see, for sufficiently small length scales these vacuum fluctuations can become important.[5] This will be explored in the next chapter.

6.8 Coherent states

The quantum oscillator has coherent states which exhibit classical behavior in macroscopic circumstances. We can create such states in the context of the electromagnetic field by applying the coherent state construction from §2.8 to each normal mode of vibration. A coherent state in the mode defined by wave vector $\vec{k}$ and polarization σ is then defined as a vector $|z_\sigma(\vec{k})\rangle$ satisfying

$$\hat{a}_\sigma|z_\sigma(\vec{k})\rangle = z_\sigma(\vec{k})|z_\sigma(\vec{k})\rangle. \tag{6.50}$$

A coherent state corresponding to a mode with wave vector $\vec{k}$ and polarization σ, all other modes in their grounds state, is built from the vacuum state by (see (2.126) in §2.9)

$$|z_\sigma(\vec{k})\rangle = e^{-|z_\sigma(\vec{k})|^2/2} e^{z_\sigma(\vec{k})\hat{a}^\dagger_\sigma(\vec{k})}|0\rangle. \tag{6.51}$$

These states will satisfy

$$\langle a_\sigma(\vec{k})\rangle = z_\sigma(\vec{k}), \tag{6.52}$$

[4] In particular, the variance of the momentum of a harmonic oscillator in a stationary state is proportional to the natural frequency of the oscillator and so it grows with frequency.

[5] Note that the electric field inside a hydrogen atom at a radius of 1 angstrom is on the order of $10^{-11} V/m$.

and will have minimum uncertainties for the real and imaginary parts of $z_\sigma(\vec{k})$. Each normal mode of vibration is an independent oscillator and one can put multiple modes into coherent states.

As time evolves the coherent states evolve according to

$$z_\sigma(\vec{k}) \quad \longrightarrow \quad z_\sigma(\vec{k},t) = e^{-i\omega(k)t} z_\sigma(\vec{k}), \tag{6.53}$$

so they remain coherent states. The expectation value of the radiation gauge vector potential in such states takes the form:

$$\langle \vec{A} \rangle(t) = \frac{1}{L^{3/2}} \sum_{\vec{k}} \sum_{\sigma=1}^{2} \sqrt{\frac{2\pi\hbar c}{k}} [z_\sigma(\vec{k})\, \vec{\varepsilon}_\sigma(\vec{k})\, e^{i\vec{k}\cdot\vec{x} - i\omega(k)t} + z^*_\sigma(\vec{k})\, \vec{\varepsilon}_\sigma(\vec{k})\, e^{-i\vec{k}\cdot\vec{x} + i\omega(k)t}]. \tag{6.54}$$

Evidently, the coherent states characterized by $z_\sigma(\vec{k})$ correspond to a mean value for the electromagnetic field which behaves as does the classical electromagnetic field defined by complex amplitudes $z_\sigma(\vec{k})$ and with minimum uncertainty in the real and imaginary parts of $z_\sigma(\vec{k})$. In this sense the coherent states can be used to understand "quasi-classical" electromagnetic phenomena.

The coherent states, while leading to familiar classical behavior for the electromagnetic field, do not have a simple interpretation in terms of photons since they involve a superposition of states with every possible photon number. Indeed, recall that coherent states of a single oscillator have a Poisson probability distribution (2.127) for energy quanta. Keeping in mind that the number of energy quanta in a given normal mode corresponds to the number of photons with the polarization and wave number of that mode, this means that the coherent states (6.51) have a Poisson distribution for the number of photons. In a coherent state for a mode characterized by the complex amplitude $z_\sigma(\vec{k})$ the probability that there are n photons is

$$P_z(n) = e^{-|z_\sigma(\vec{k})|^2} \frac{|z_\sigma(\vec{k})|^{2n}}{n!}. \tag{6.55}$$

In such a coherent state, the average value of the number of photons in the normal mode labeled by $(\sigma, \vec{k})$ is then given by

$$\langle N_\sigma(\vec{k}) \rangle = \sum_{n=0}^{\infty} n P_z(n) = |z_\sigma(\vec{k})|^2, \tag{6.56}$$

so the absolute value-squared of the coherent state eigenvalue, representing the squared modulus of the classical complex amplitude, is the mean number of photons in the coherent state $|z_\sigma(\vec{k})\rangle$. This is how you can connect the intensity of a classical electromagnetic wave with the mean number of photons in the quantum state it approximates.[6]

6.9 Photon interference

When considering the behavior of photons in an interference experiment (*e.g.,* interferometer or double slit) Dirac has famously reasoned that "Each photon ... interferes only with itself." [1]. I cannot improve upon his reasoning nor his exposition of it; I invite you to read it for yourself[7] Here

[6] It also shows you that two common mental images for a photon are untenable. One erroneous mental image is that a photon is akin to a classical electromagnetic plane wave. Another erroneous mental image is that a photon is some localized wave packet of the otherwise classical electromagnetic field. Classical electromagnetic plane waves and wave packets corresponds to coherent states of the quantum electromagnetic field. For such electromagnetic field configurations there is a non-zero probability for any number of photons – even zero!

[7] This quote appears in the introduction to reference [1], where the need for a quantum theory is examined. This introduction is very clear and still relevant today; it is well worth your time to read it in its entirety.

I would simply like to sketch how the formalism of quantum field theory implements this idea.

In a classical interference experiment a beam of light is split into two parts and then recombined in a region where the intensity can be measured as a function of position. The interference manifests itself in regions of maxima and minima for the observed intensity. This can be understood from Maxwell's equations where the two parts of the split beam are described by solutions $\vec{A}_1$ and $\vec{A}_2$ to the Maxwell equations. For example, one can model the "slits" as very small pinholes each of which produces a spherical wave of the form

$$\vec{A}_{1,2} = a\frac{\sin\theta}{kr_{1,2}}\left(\sin(kr_{1,2}-\omega t)+\frac{\cos(kr_{1,2}-\omega t)}{kr_{1,2}}\right)\hat{\phi}, \tag{6.57}$$

where I am using spherical coordinates centered at the pinhole; a is a constant, $\omega = kc$, and $\hat{\phi}$ is the unit vector in the azimuthal direction. The total electromagnetic field is the superposition $\vec{A} = \vec{A}_1 + \vec{A}_2$. In the region where intensities are measured the beams of light have combined – the supports of $\vec{A}_1$ and $\vec{A}_2$ overlap. The intensity is proportional to the square of the electric field

$$\vec{E} = -\frac{1}{c}\left(\dot{\vec{A}}_1 + \dot{\vec{A}}_2\right), \quad E^2 = \frac{1}{c^2}\left(\dot{A}_1^2 + \dot{A}_2^2 + 2\dot{\vec{A}}_1\cdot\dot{\vec{A}}_2\right). \tag{6.58}$$

The first two terms in E^2 provide the intensity of each of the component beams of light, while the third term brings the interference into play.

As we have discussed already, macroscopic/classical electromagnetic phenomena involve states of the quantum electromagnetic field which have non-vanishing probabilities for any number of photons. Nowadays it is common to consider situations where one is making measurements on states of the quantum electromagnetic field which consist of a small number of photons – even a single photon. In particular, one can perform interference experiments in which the beam of light has been reduced in intensity so that individual photons can be studied. While it is difficult to guarantee that such experiments always work with one photon at a time, it is quite easy to see what interference phenomena are predicted by quantum field theory when this is the case.

The simplest model for photon interference phenomena comes by supposing that one has a way of creating a photon in one of two modes of the electromagnetic field, which we can label 1 and 2. We can use an occupation number basis which includes these two modes, and we write $|1\rangle$, $|2\rangle$ for the Fock states where a single photon occupies mode 1 or mode 2, respectively. If one is thinking about a double-slit experiment, the two photon modes correspond to electromagnetic waves coming from one or the other slit. If one is thinking of an interferometer, the two modes correspond to waves propagating in one arm or the other. Mathematically, these modes correspond to the solutions to the Maxwell equations $\vec{A}_1$ and $\vec{A}_2$, *e.g.,* (6.57), appearing in the classical description summarized above. The change of basis from the plane wave modes to the modes of interest for the interference problem is an instance of the general formalism described in §4.5. In particular, while the observable characteristics of photons associated to the new basis of modes will be different, the total number of photons (if it is statistically certain) is the same in any basis. A photon which is in a superposition of the two modes is described by the family of (Fock) state vectors

$$|\theta,\phi\rangle = \cos\theta|1\rangle + e^{i\phi}\sin\theta|2\rangle, \tag{6.59}$$

where θ and ϕ are determined by the details of the experimental preparation procedure. Notice that this is the most general normalized linear combination of the two orthogonal unit vectors $|1\rangle$ and $|2\rangle$. If we introduce a photon detector, what will it detect? To answer that question we have to think about how a photon detector works.

Most photon detectors involve absorption of a photon by an atom (or some other bound state of electrical charges), leaving it in an ionized state which can be identified via a macroscopic amplification process, *e.g.,* a cascade of electrons producing a measurable current. As we shall examine in some detail in the next chapter, for a simple hydrogenic atom and to first order in perturbation theory,[8] the probability for this process to occur at time t (and hence the detection rate) will be proportional to:

$$\mathcal{I} \equiv |\langle 0|\hat{\vec{E}}(\vec{x}_0, t)|\theta, \phi\rangle|^2, \tag{6.60}$$

where $\vec{x}_0$ is the position of the atom of interest. The matrix element of interest takes the form

$$\langle 0|\hat{\vec{E}}(\vec{x}_0, t)|\theta, \phi\rangle = \vec{u}_1(\vec{x}_0, t)\cos\theta + \vec{u}_2(\vec{x}_0, t)e^{i\phi}\sin\theta, \tag{6.61}$$

where $\vec{u}_1$ and $\vec{u}_2$ are the solutions to the Maxwell equations describing the two modes of interest (*e.g.*, 6.57)). We then get

$$\mathcal{I} = |u_1(\vec{x}_0, t)|^2\cos^2\theta + |u_2(\vec{x}_0, t)|^2\sin^2\theta + \cos(2\theta)\Re\left(e^{i\phi}\vec{u}_2(\vec{x}_0, t)\cdot\vec{u}_1^*(\vec{x}_0, t)\right). \tag{6.62}$$

Keep in mind that θ and ϕ are determined by the experimental preparation procedure. For example, a simple model of the two slit experiment prepares the photon in the state with $\theta = \pi/4$, $\phi = 0$. In any case, the interference effects come from the relative phases of $\vec{u}_1$ and $\vec{u}_2$ as a function of detection location $\vec{x}_0$. If we write (at location $\vec{x}_0$ and time t)

$$\vec{u} = \vec{R}e^{i\alpha}, \quad R \text{ real}, \tag{6.63}$$

then

$$\Re\left(e^{i\phi}\vec{u}_2(\vec{x}_0, t)\cdot\vec{u}_1^*(\vec{x}_0, t)\right) = (\vec{R}_1\cdot\vec{R}_2)\cos(\phi + \alpha_1 - \alpha_2) \tag{6.64}$$

This is the same interference formula one gets from the classical description. As the detector location varies the relative phase of the two modes varies producing an interference pattern of photon counts.

6.10 Limitations on the particle interpretation

You may have noticed that in all of our discussion of the physical characteristics of photons the notion of *where* is the photon has been carefully avoided. There is a good reason for this. To speak about *the* location of a photon – or any other particle, for that matter – requires constructing observables and states which provide a way to intrinsically "localize" a photon to a given spatial region. This is not a trivial task since one must construct this notion of location in a way which is compatible with the principles of the special theory of relativity. For a massive particle like the electron, given a few more or less self-evident hypotheses, it is possible to *uniquely* construct such observables and states independently of any new structures (*e.g.,* some new degrees of freedom to represent a source or a measuring apparatus) [6], [7]. One then interprets this state of affairs as indicating that the electron has an intrinsic notion of "location" as part of its "reality".[9] However, the same desiderata which lead to a unique notion of position for an electron lead to a non-existence result for the position operator for a photon. Indeed, for any massless particle with non-zero helicity

[8] Using Fermi's Golden Rule and the electric dipole approximation.

[9] Even for an electron, however, there are interesting limits on its localizability in its proper relativistic description.

this difficulty will occur, as shown by Newton-Wigner [6] and Wightman [7]. See also a more accessible proof due to Jordan [8].

How does one prove something like this? Well, it's a pretty fancy bit of mathematical physics — too fancy for this humble text — but the idea is as follows. One assumes that the states of a quantum system behave properly under changes of inertial reference frame in accord with the tenets of the special theory of relativity. This means that one can characterize "particles" according to their rest mass and spin – or helicity if the rest mass is zero. One then searches for a position operator, defined to be a trio of self-adjoint operators that behave like position with respect to spatial translations, spatial rotations, and time reversal (in any given inertial reference frame). Explicitly, the would-be position operator should rotate as a spatial vector, translate like a spatial position, and be invariant under time reversal. Finally, the three operators comprising the components of the position vector are required to be compatible – their operator representatives are required to commute so that one has a complete set of (generalized) eigenvectors of the position. For massive particles or for massless particles with no helicity, the position operator is *uniquely* determined by these criteria. But for massless particles with non-zero helicity it can be shown there is no such operator! If the putative position operator is allowed to have non-commutative components, so that photons can't be localized, strictly speaking, then the "position" operator and its associated notion of localization is not unique. Different choices of "position" then lead to different notions of photon localization. Physically, this means the "location" of a photon can be expected to be contingent on the experimental means used to measure its location.

It appears then that photons are not endowed with an intrinsic observable corresponding to position. Without some additional extrinsic structures (or a more limited notion of position) the notion of a photon as a point particle like, say, the electron is limited at best. Of course, if the photon is viewed as fundamentally a quantum normal mode of the electromagnetic field then this state of affairs is perhaps not quite so striking.

6.11 Problems

1. Show that the energy, momentum and helicity operators (6.2), (6.9), (6.19) all commute. (This means there is a basis of states where these observables have statistically sharp values.)

2. Consider the photon state (6.29). Show that this state does contain a single photon with certainty. What is the probability for finding helicity ± 1?

3. Verify (6.43).

4. Show that the electric and magnetic field are not *compatible*, that is, their operator representatives do not commute. (This indicates that the electric and magnetic fields will satisfy an uncertainty relation analogous to the position and momentum of a particle.

5. You are making some tasty microwave popcorn (with extra butter). Estimate how much electromagnetic energy is in the ostensibly classical microwaves; estimate the magnitude of the electromagnetic field and compare it with the uncertainty we calculated. Estimate how many photons there are in the microwave electromagnetic field.

6. Calculate the variance of the average magnetic field in the vacuum state as we did for the electric field and show that similar results occur.

7. Show that the energy operator defined by

$$\hat{\mathcal{E}} = \frac{1}{8\pi} \int_{\text{all space}} d^3x \left(\hat{E}^2 + \hat{B}^2 \right)$$

differs from the Hamiltonian (6.2) by an infinite(!) term proportional to the identity operator. This infinite "zero point energy" is normally deemed unobservable, but the difference between such zero point energies can have observable effects. (Search the internet for the *Casimir effect*.)

8. It is sometimes proposed that the vector potential $\vec{A}(\vec{x}, t)$ in the Coulomb gauge ($\nabla \cdot \vec{A} = 0$) can be used as a position wave function at time t for the photon, and we can then define the position operator for the photon (and probabilities for its position) in the usual way, $\hat{\vec{x}}\vec{A} = \vec{x}\vec{A}$. Can you find a flaw in this construction? (*Hint:* The way this is supposed to work is that the wave function is a 3-component column vector, much like the wave function for a spin 1/2 particle is a 2-component column vector. One needs three components because the photon has spin-1. The divergence-free condition is needed to remove the helicity zero state, which cannot occur in a massless particle.)

9. Suppose that each normal mode of vibration of the quantum electromagnetic field is in a coherent state (see Chapter 2). Consider the expectation values of $\vec{E}$ and $\vec{B}$ and show that they take the classical form of solutions to the source-free Maxwell equations. How would you interpret these states in terms of photons?

10. Prove the formula (6.56) for the mean number of photons in a coherent state.

Chapter 7

Spontaneous Emission

In this chapter the aim is to analyze the process of spontaneous emission of photons from atoms using the quantum electromagnetic field of the previous chapter. For simplicity we will focus on the simplest kind of atom – the hydrogenic atom. In the following presentation I have borrowed heavily from the treatments of Merzbacher [2] and Sakurai [3].

7.1 Hydrogenic atoms

A *hydrogenic* atom is a model of an atom that is useful when one only needs a single electron to analyze the physical processes of interest. Of course, a hydrogen atom fits the bill, and so will ions with a single electron. Atoms with a single valence electron may also behave this way for many purposes. Our model for a hydrogenic atom will be a single negatively charged particle of mass m and negative charge $-q < 0$ bound by a Coulomb field to a nucleus (fixed in space) with positive charge Zq. We will ignore the effects of the electron and/or nuclear spin. The Hilbert space is then the set of square-integrable functions in $\mathbf{R}^3$ and the Hamiltonian is

$$\hat{H}_0 = \frac{\hat{p}^2}{2m} - Z\frac{q^2}{\hat{r}}, \tag{7.1}$$

so that

$$\hat{H}_0\psi(\vec{x}) = -\frac{\hbar^2}{2m}\nabla^2\psi(\vec{x}) - \frac{Zq^2}{r}\psi(\vec{x}). \tag{7.2}$$

The atomic energy levels come from the bound states – the discrete part of the spectrum of H_0. The energy eigenvalues are

$$E_n = -\frac{Z^2q^2}{2a}\frac{1}{n^2}, \quad n = 1, 2, 3, \ldots, \tag{7.3}$$

where a is the *Bohr radius*

$$a = \frac{\hbar^2}{mq^2}. \tag{7.4}$$

The eigenfunctions are, in spherical polar coordinates,[1]

$$\psi_{nlm}(r, \theta, \phi) = R_{nl}(r)Y_l^m(\theta, \phi), \tag{7.5}$$

[1] Here θ is the polar angle and ϕ is the azimuthal angle.

where Y_l^m are the spherical harmonics (normalized on the unit sphere) and

$$R_{nl} = -\sqrt{\left(\frac{2Z}{na}\right)^3 \frac{(n-l-1)!}{2n[(n+l)!]^3}} e^{-\rho/2} \rho^l L_{n+l}^{2l+1}(\rho), \tag{7.6}$$

where

$$\rho = \frac{2Z}{na} r, \quad n \geq l+1, \tag{7.7}$$

and $L_{n+l}^{2l+1}(\rho)$ is an associated Laguerre polynomial:

$$L_k^s(\rho) = \frac{d^s}{d\rho^s}\left(e^\rho \frac{d^k}{d\rho^k}(\rho^k e^{-\rho})\right). \tag{7.8}$$

Yes, it all looks very complicated. But we are fortunate to have explicit formulas. Indeed, if you add another electron to the atom (to get helium) there are no such formulas available, and you must resort to approximation methods to analyze the stationary states.

To use these formulas for the bound state spectrum you should first pick a value for n. This fixes the energy of interest according to (7.3. Then you can pick any value for l such that $l = 0, 1, 2, \ldots, n-1$. This fixes the magnitude of the angular momentum of the state to be $\hbar\sqrt{l(l+1)}$. Then you pick any m from the set $m = -l, -l+1, \ldots, l-1, l$. This fixes the value of one component of the angular momentum vector to be $m\hbar$. (This component is the z component in the spherical polar coordinate system above.) The state is uniquely determined by these three observables.[2] The ground state of the atom arises when $n = 1$, $l = 0$, $m = 0$. Its wave function is

$$\psi_{100} = \sqrt{\frac{Z^3}{\pi a^3}} e^{-Zr/a}. \tag{7.9}$$

A first excited state we shall be interested in is

$$\psi_{210} = \sqrt{\frac{Z^3}{32\pi a^3}} \frac{Zr}{a} \cos\theta e^{-Zr/2a}. \tag{7.10}$$

The states are sometimes labeled using a traditional spectroscopic notation. In this labeling scheme the angular momentum is denoted by a letter, $S, P, D, F, \ldots$ corresponding to $l = 0, 1, 2, 3, \ldots$ while the energy level is denoted by its "principal quantum number" n. For example, the $1S$ state is the ground state; the $2P$ state is a first excited state and has $l = 1$.

7.2 The hydrogenic atom coupled to photons

We have spent considerable time building the quantum theory of the electromagnetic field and its photon manifestations using the Fock space $\mathcal{F}$. We have just reviewed some of the quantum theory of a hydrogenic atom and its space of quantum states $\mathcal{H}$.[3] Now I would like to put these two subsystems together to describe a quantum system consisting of a hydrogenic atom and the

[2] The fact that the energy does not depend m is due to rotational symmetry of the Hamiltonian. The fact that the energy doesn't depend upon l is a special feature of hydrogenic atoms. It stems from a "hidden symmetry" of the system, which only arises for potentials proportional to $1/r$ or proportional to r^2.

[3] Of course, we only explicitly treated the bound states.

quantum electromagnetic field. The ultimate model for this requires a fully relativistic field theoretic treatment of all the constituents – this is the theory of quantum electrodynamics or, to go all the way, the standard model of particle physics. If we just want to understand phenomena like spontaneous emission we can use a more humble model in which we combine the atomic model reviewed above with the quantum electromagnetic field of the previous chapter. We then introduce a tried and true interaction between the systems to get non-trivial – and experimentally observed – phenomena. We do this as follows.

The Hilbert space of states of the combined system, consisting of the hydrogenic atom and the quantum electromagnetic field, is taken to be the tensor product $\mathcal{H} \otimes \mathcal{F}$. For our purposes, a very useful basis is induced by the respective energy eigenvectors:

$$|E, \mathbf{n}\rangle \equiv |E\rangle \otimes |\mathbf{n}\rangle, \tag{7.11}$$

where $|E\rangle$ are the basis of (generalized) eigenvectors of $\hat{H}_0$ in (7.2), whose bound state eigenvectors are explicitly displayed in (7.5), and $|\mathbf{n}\rangle$ are the Fock states defined by occupation number $\mathbf{n} = n(\vec{k}, \sigma)$ for each normal mode labeled by wave vector $\vec{k}$ and helicity $\sigma = \pm$. The vectors (7.11) constitute a basis of states where the atomic energy level, and the various photon energies, momenta, and helicities are known with certainty. In particular, the usual scenario where the electron is in some atomic energy level and there are no photons present is described by states of the form $|E_k\rangle \otimes |0\rangle$. The basis (7.11) is a basis of eigenvectors of a Hamiltonian $\hat{\mathbf{H}}_0$ built by adding the Hamiltonians for the atom and for the electromagnetic field:

$$\hat{\mathbf{H}}_0 = \left(\frac{\hat{p}^2}{2m} - \frac{Zq^2}{\hat{r}} \right) \otimes \hat{1} + \hat{1} \otimes \left(\sum_{\vec{k},\sigma} \hbar kc \; \hat{a}^\dagger_\sigma(\vec{k}) \hat{a}_\sigma(\vec{k}) \right), \tag{7.12}$$

$$\hat{\mathbf{H}}_0 |E, \mathbf{n}\rangle = (E + \mathcal{E})|E, \mathbf{n}\rangle, \quad \mathcal{E} = \sum_{\vec{k},\sigma} n(\vec{k}, \sigma) \hbar kc. \tag{7.13}$$

With dynamics defined by $\hat{\mathbf{H}}_0$ the vectors (7.11) represent stationary states and will not exhibit any time dependent behavior, *e.g.,* transitions between states. To get such behavior we must add new terms to the Hamiltonian which represent the energy of interaction between the atomic electron and the quantum electromagnetic field itself. To be sure, the potential energy term in (7.12) represents the electromagnetic interaction with the nucleus – indeed, it will be by far the dominant interaction in our investigation. Our approximation strategy is to take the Coulomb interaction of the nucleus as given and study the interaction of the atomic electron with small quantum fluctuations of the electromagnetic field about the Coulomb field. This is adequate to understand processes like spontaneous emission. It also jibes with the discussion of the influence of prescribed sources in §5.5 in classical electrodynamics. What I mean is this: classically, the electromagnetic field experienced by an atomic electron would (in the Coulomb gauge) involve (i) the scalar potential taking its usual Coulomb form, and (ii) the vector potential – solving (5.64) – representing any electromagnetic radiation which might be present. Spontaneous emission occurs in the absence of any electromagnetic radiation. Classically the absence of radiation would be modeled by $\vec{A} = 0$, but quantum mechanically things are more interesting, as we have seen in §6.7.

Our model of the hydrogenic electron already includes the Coulomb interaction with the nucleus. We incorporate the quantum electromagnetic field by turning (5.111) into an operator on $\mathcal{H} \otimes \mathcal{F}$ with ϕ providing the atomic Coulomb field and $\vec{A}$ providing the quantum electromagnetic fluctuations.

The Hamiltonian we shall start from is then

$$\hat{\mathbf{H}} = \frac{1}{2m}\left[\hat{\vec{p}} - \frac{q}{c}\hat{\vec{A}}(\hat{\vec{x}})\right]^2 - \frac{Zq^2}{\hat{r}} + \sum_{\vec{k},\sigma} \hbar kc\, \hat{a}^\dagger_\sigma(\vec{k})\hat{a}_\sigma(\vec{k}) \tag{7.14}$$

$$= \hat{\mathbf{H}}_0 - \frac{q}{mc}\hat{\vec{A}}(\hat{\vec{x}})\cdot\hat{\vec{p}} + \frac{q^2}{2mc^2}\hat{A}^2(\hat{\vec{x}}). \tag{7.15}$$

To get the second term in this formula I used the fact that the vector potential has vanishing divergence in the radiation gauge. We will be using the Schrödinger picture, so we get $\hat{\vec{A}}(\hat{\vec{x}})$ by setting $t = 0$ in (6.35). Notice also that now we are treating the quantum vector potential as an operator on $\mathcal{H}\otimes\mathcal{F}$. In particular the operator $\hat{\vec{A}}(\hat{\vec{x}})$ is defined via the formula (6.35) but with the substitution $\vec{x} \to \hat{\vec{x}}$ since $\vec{x}$ in the Hamiltonian is meant to represent the position of the electron. The position $\hat{\vec{x}}$ in the vector potential acts on $\mathcal{H}$ and the creation/annihilation operators act on $\mathcal{F}$ as usual with a tensor product.

The presence of the last two terms in (7.15) means that the basis (7.11) no longer consists of stationary states. If the initial state of the system corresponds to the electron being in one of the atomic energy levels with no photons present, after some time has elapsed there may be a non-zero probability for the electron to be in a different energy level with some photons present. To calculate this probability the strategy will be to approximate the effect of the terms involving $\hat{\vec{A}}$ under the assumption that such effects are small, as we empirically know them to be.

7.3 Perturbation theory. Fermi's Golden Rule

In order to study the atomic transitions induced by the quantum electromagnetic field we need some basic results from time-dependent perturbation theory and, in particular, "Fermi's Golden Rule" for transitions. These are standard results from quantum mechanics and we could simply take them as given, but I would like to make a small detour to review the derivation of these results since I think it is very interesting to see precisely which approximations yield the usual lore of atomic transitions.

Let us denote the interaction part of the Hamiltonian by the operator $\hat{V}$,

$$\hat{\mathbf{H}} = \hat{\mathbf{H}}_0 + \hat{V}. \tag{7.16}$$

One result from time-dependent perturbation I *will* take as given is that, assuming the effects of $\hat{V}$ are small (and assuming the operator $\hat{V}$ does not depend upon time), if the system starts out in an eigenstate $|i\rangle$ of $\hat{\mathbf{H}}_0$ with energy E_i, the probability for making a transition to a *different* eigenstate $|f\rangle$ of $\hat{\mathbf{H}}_0$ with energy E_f after elapsed time t is

$$P(i\to f) = \frac{4|\langle f|V|i\rangle|^2}{(E_f - E_i)^2}\sin^2\left\{\frac{(E_f - E_i)t}{2\hbar}\right\}, \quad |i\rangle \neq e^{i\alpha}|f\rangle. \tag{7.17}$$

Here it is assumed that the unperturbed stationary states are part of the discrete spectrum; we will relax this assumption in a moment. For now, you can see that the effect of the perturbation represented by $\hat{V}$ is to induce transitions with a sinusoidally varying probability provided $E_f \neq E_i$.

Notice that the time scale Δt for the sinusoidal function to take an appreciable value depends upon the energy difference $\Delta E = (E_f - E_i)$ according to

$$\Delta t \Delta E \sim \hbar, \tag{7.18}$$

in accord with the time-energy uncertainty principle. As the final (unperturbed) energy differs more and more from the initial energy the transition probability is suppressed by the factor in front of the sine function. If the eigenvalue E_i is degenerate it is possible to have $E_f = E_i$. While these so-called "energy conserving transitions" (if any) will dominate, there is a relatively small but non-zero probability for "energy non-conserving transitions". This "conserving" and "non-conserving" terminology is helpful, but you should keep in mind that it is just a way of speaking about the transitions in terms of the unperturbed system. The true energy of the system is defined by eigenvalues of $\hat{\mathbf{H}}$ and the probability distribution for eigenvalues of $\hat{\mathbf{H}}$ is conserved. In particular, if the system starts off in a eigenstate of $\hat{\mathbf{H}}$ it remains in that eigenstate for all time. For the energy conserving transitions (if any) the sinusoidal behavior in time of the transition probability is an illusion because of a resonance phenomenon. For $E_f = E_i$,

$$P(i \to f) = \frac{|\langle f|V|i\rangle|^2}{\hbar^2} t^2 \quad |i\rangle \neq e^{i\alpha}|f\rangle, \quad E_f = E_i, \tag{7.19}$$

and you can see that the probability for energy conserving transitions grows quadratically in time. Since the perturbation expansion is valid only for sufficiently small transition probabilities, the perturbative formula is only valid for sufficiently short elapsed time in the energy conserving case, depending upon the size of the transition matrix element.

We have seen that for transitions between unperturbed stationary states associated with the discrete spectrum of $\hat{\mathbf{H}}_0$ the effect of the perturbation is to cause a periodic transition to/from the initial state, with "energy conserving" transitions (if there are any available) becoming dominant after a sufficiently large time interval. Here "large time" means that the elapsed time is much larger than the period of oscillation of the transition probability for energy-non-conserving transitions,

$$t >> T \equiv \frac{2\pi\hbar}{|E_f - E_i|}. \tag{7.20}$$

Note that the typical energy scale for atomic structure is on the order of electron-volts. This translates into a typical time scale $T \sim 10^{-15} s$, so "large time" in the sense of (7.20) is often a very good approximation for atomic systems, even while the time can still be sufficiently short to use first-order perturbation theory.

In the foregoing we have been assuming the final state is an energy eigenstate coming from the discrete part of the energy spectrum. In many important examples the final state of interest lies in a continuum (at least approximately). This will the case in the situation of interest to us since the final state will have a photon whose wave vector varies more or less continuously (for large enough L). In this case qualitatively new features emerge: the periodic transitions to/from the initial state are replaced with irreversible transitions away from the initial state. We shall see that the transition probability at "large times" still favors energy conserving transitions, but it will only grow linearly with time because the width of the probability distribution about such transitions is becoming narrower with time.

We can see this by supposing the final state is part of a continuum of states labeled by the continuous (unperturbed) energy E and possible some other observables, which I will generically

denote by ζ. The transition probability is now a probability density (at least in energy space) and is to be integrated/summed over some range in the (E, ζ) variables to get a transition probability. Assuming $t >> T$, the probability density can be computed via the representations of the delta function $\delta(x)$, where x is a continuous variable:

$$\lim_{b\to\infty} \frac{\sin^2(bx)}{bx^2} = \pi\delta(x), \quad \delta(ax) = \frac{1}{|a|}\delta(x).$$

Then we have for the transition *density* (exercise)

$$P(i \to f) \approx \frac{2\pi}{\hbar}|V_{fi}(E,\zeta)|^2\delta(E - E_i)\, t, \quad t >> T.$$

Here I have defined

$$V_{fi} \equiv \langle f|\hat{V}|i\rangle. \tag{7.21}$$

As you can see, at sufficiently late times and with a quasi-continuum of final states, the *transition rate density* $\frac{dP}{dt}$ is non-negligible only for energy conserving transitions and – in contrast to transitions to states in the discrete spectrum – is constant in time:

$$\frac{d}{dt}P(i \to f) \approx \frac{2\pi}{\hbar}|V_{fi}(E,\zeta)|^2\delta(E - E_i), \quad t >> T.$$

a result which is one version of *Fermi's Golden Rule*.

In situations where there is a continuum (or near continuum) of final energies we are interested in the transition rate to final states with energy in some range $\mathcal{E}$. We then need to integrate the transition rate density over $\mathcal{E}$ and sum/integrate over the remaining variables ζ to get a transition rate. The result will involve a "density of states" factor $\rho(E, \zeta)$ which relates an integral over energy to an integral over states. The quantity $\rho(E, \zeta)dE$ is the number of states with energies between E and $E + dE$. In general the density of states will also depend upon the variables ζ, which are taken to be discrete, although it is no problem if they are continuous. The transition rate w from the initial state $|i\rangle$ into states $|f\rangle$ with energy $E \in \mathcal{E}$ is then expressible in the form

$$w = \sum_{\zeta} \int_{\mathcal{E}} dE\, \rho(E,\zeta)\frac{2\pi}{\hbar}|V_{fi}|^2\delta(E - E_i) = \frac{2\pi}{\hbar}\sum_{\zeta} \rho(E_i,\zeta)|V_{fi}|^2.$$

This is probably the most commonly used version of Fermi's Golden Rule.

7.4 Spontaneous emission

7.4.1 First-order approximation

To approximately incorporate the quantum electromagnetic field effects on the atom, we use the results from time-dependent perturbation theory which were reviewed in the previous section. Let us focus on transitions in which the initial state has the atom in an unperturbed stationary state with energy E_j and with no photons present, and in which the final state has the atom in an unperturbed stationary state with energy E_n with a single photon present. This is the simplest

spontaneous emission process.[4] The matrix elements of interest is then

$$V_{fi} = \langle E_n, 1_{\vec{k},\sigma}| \left[-\frac{q}{mc}\hat{\vec{A}}(\hat{\vec{x}}) \cdot \hat{\vec{p}} + \frac{q^2}{2mc^2}\hat{A}^2(\hat{\vec{x}}) \right] |E_j, 0\rangle, \tag{7.22}$$

where

$$|E_j, 0\rangle = |E_j\rangle \otimes |0\rangle, \quad |E_n, 1_{\vec{k},\sigma}\rangle = |E_n\rangle \otimes |1_{\vec{k},\sigma}\rangle = |E_n\rangle \otimes \left(a^\dagger_\sigma(\vec{k})|0\rangle \right), \tag{7.23}$$

and where *a priori* all wave vectors $\vec{k}$ and polarizations σ have to be considered since the probability for spontaneous emission is obtained by summing over all final photon states. The first term in (7.22) is linear in creation and annihilation operators while the second term is quadratic. For the single photon process of interest only the creation operators in the first term contribute. (The quadratic term yields transitions involving emission/absorption of 2 photons.) Thus the matrix element becomes:

$$V_{fi} = -\frac{q}{mc}\sqrt{\frac{2\pi\hbar c}{kL^3}} \langle E_n | e^{-i\vec{k}\cdot\vec{x}} \vec{\varepsilon}_\sigma(\vec{k}) \cdot \hat{\vec{p}} | E_j \rangle. \tag{7.24}$$

Notice that the photon states are now out of this formula, which only refers to the unperturbed atomic energy levels.

For large enough L the final photon states form a quasi-continuum (because the wave vectors do) and we are in the domain of Fermi's Golden Rule. Consequently, assuming $t >> 10^{-15}\, s$, transitions occur (predominately) to final states with the same unperturbed energy as the initial state,

$$E_n + \hbar\omega(k) = E_j, \tag{7.25}$$

with a transition rate $w_{\sigma,\vec{k}}$ which is constant in time and given by

$$w_{\sigma,\vec{k}} = \frac{2\pi}{\hbar} |V_{jn}(\vec{k}, \sigma)|^2 \rho(E_n + \hbar\omega(k)), \tag{7.26}$$

where $\rho(E)$ is the density of final states at energy E. To calculate the density of states we note that each state is determined by its helicity state ± 1 and a triple of integers

$$(n_x, n_y, n_z) = \frac{L}{2\pi}\vec{k}. \tag{7.27}$$

For each helicity state of the photon, summing over states in the quasi-continuum around the final energy can (for large enough L) be performed by integrating over $\vec{k}$ with a factor of $(L/2\pi)^3$. In spherical polar coordinates in $\vec{k}$ space the integration measure for summing over states is then

$$\left(\frac{L}{2\pi}\right)^3 k^2 dk d\Omega = \frac{\omega(k)^2 L^3}{8\pi^3\hbar c^3} d\Omega dE. \tag{7.28}$$

This means the photon density of states for each helicity with wave vector in the solid angle $d\Omega$ (in k-space) is given by

$$\rho(E) = \frac{\omega(k)^2 L^3}{8\pi^3\hbar c^3} d\Omega. \tag{7.29}$$

[4] Can you explain why there is no spontaneous absorption?

(In the notation of the previous section, the energy of interest is $\hbar kc$ and the variables generically denoted by ζ correspond to the helicity and the direction of the wave vector.) Therefore the transition rate is

$$w_{\sigma,\vec{k}} = \frac{\alpha}{2\pi(mc)^2}\omega(k)\left|\langle E_n|e^{-i\vec{k}\cdot\hat{\vec{x}}}\vec{\epsilon}_\sigma(\vec{k})\cdot\hat{\vec{p}}|E_j\rangle\right|^2 d\Omega, \tag{7.30}$$

where $\alpha = q^2/\hbar c$ (with q being the magnitude of the electron charge) is the fine structure constant. Notice that this result is independent of the precise value of L; we only need it to be sufficiently large.[5]

7.4.2 Electric dipole transitions

It only remains to deal with the matrix element in (7.30). Fortunately for me, this matrix element is frequently treated in courses on quantum mechanics and a standard approximation to it is available. The idea is as follows. Because the transition is "energy conserving", the wavelength λ of the emitted photon is related to the change in atomic energy ΔE in the usual way:

$$\lambda = \frac{2\pi\hbar c}{\Delta E}, \quad k = \frac{2\pi}{\lambda} = \frac{\Delta E}{\hbar c} \tag{7.31}$$

The wave number k of the emitted photon is on the order of $\sim 10^7\, m^{-1}$. On the other hand, the position space wave functions being integrated in the matrix element are non-negligible only out to the Bohr radius or so, say $10^{-10}\, m$. Consequently, the exponent in (7.30) can be neglected in a first approximation:

$$w_{\sigma,\vec{k}} \approx \frac{\alpha}{2\pi(mc)^2}\omega(k)\left|\langle E_n|\vec{\epsilon}_\sigma(\vec{k})\cdot\hat{\vec{p}}|E_j\rangle\right|^2 d\Omega, \tag{7.32}$$

It is now possible to calculate the matrix element explicitly, but it is useful to first put the result in a more illuminating form via the identity

$$\hat{\vec{p}} = \frac{m}{i\hbar}[\hat{\vec{x}}, \hat{\mathbf{H}}_0]. \tag{7.33}$$

Replacing the momentum with the commutator, and then taking advantage of the fact that the matrix elements are with respect to eigenvectors of $\hat{\mathbf{H}}_0$, we get

$$w_{\sigma,\vec{k}} \approx \frac{\alpha}{2\pi c^2}\,\omega(k)^3\left|\langle E_n|\vec{\epsilon}_\sigma(\vec{k})\cdot\hat{\vec{x}}|E_j\rangle\right|^2 d\Omega. \tag{7.34}$$

The operator $q\hat{\vec{x}}$ represents the electric dipole moment of the atom and the dot-product with the polarization vector is proportional to the energy of an electric dipole in a uniform electric field along $\vec{\epsilon}_\sigma(\vec{k})$. For this reason the transitions which are considered in this approximation are known as *electric dipole transitions*. Using the theory of angular momentum, it is well-known that the electric dipole matrix element is non-vanishing only when certain *selection rules* for the difference in angular momentum values, Δl, Δm, are satisfied. The electric dipole selection rules are

$$\Delta l = \pm 1, \quad \Delta m = 0, \pm 1 \tag{7.35}$$

[5] Since the typical photon wavelengths which matter for this transition are $\sim 10^{-7}\, m$, any macroscopic box will be sufficiently large.

The relation $\Delta m = 0$, comes from the term involving the z component of the polarization vector $\vec{\epsilon}_\sigma$, while it is the x-y components of the polarization which feature when $\Delta m = \pm 1$.

It is worth pausing to think about this selection rule in terms of conservation of angular momentum.[6] The Hamiltonian for the system – atom plus photons – is rotationally invariant and therefore the total angular momentum is conserved. This angular momentum comes from the atomic orbital angular momentum and the photon's helicity, which only contributes angular momentum parallel or anti-parallel to its motion. The change in the total angular momentum of the atom, $\Delta l = \pm 1$, is consistent with conservation of total angular momentum and the spin-1 nature of the photon. (Notice that it is possible for the angular momentum of the atom to *increase* by emission of a photon.) The z component of the total angular momentum is just the sum of the z components of the individual angular momenta of the atom and photon. If the photon is emitted along the z axis then its polarization vector is in the x-y plane, whence the selection rule $\Delta m = \pm 1$ applies, which jibes with the idea that the component of the photon's intrinsic angular momentum along the direction of propagation of the photon (in a momentum eigenstate) is $\pm\hbar$. If the photon has polarization only along z then it is propagating in the x-y plane and has no intrinsic angular momentum along z, consistent with the selection rule $\Delta m = 0$.

To get the transition rate in final form we sum over all final photon states. This involves a discrete sum over polarizations and an integral over wave vector directions. We take care of this as follows. The absolute value of each component of the matrix element of $\hat{\vec{x}}$ defines a vector in $\mathbf{R}^3$ which I will denote by $\vec{x}_{nj}$,

$$\vec{x}_{nj} \equiv |\langle E_n|\hat{\vec{x}}|E_j\rangle|. \tag{7.36}$$

Recall that the vectors $(\vec{\epsilon}_1(\vec{k}), \vec{\epsilon}_2(\vec{k}), \vec{k}/k)$ form an oriented (right-handed) orthonormal basis in $\mathbf{R}^3$. Let θ be the angle between $\vec{k}$ and $\vec{x}_{nj}$. Imagine projecting $\vec{x}_{nj}$ into the plane orthogonal to $\vec{k}$. Let ϕ be the angle of that projection with respect to $\vec{\epsilon}_1$. We can then write

$$\vec{x}_{nj} = x_{nj}\left(\cos\theta\,\frac{\vec{k}}{k} + \sin\theta\cos\phi\,\vec{\epsilon}_1 + \sin\theta\sin\phi\,\vec{\epsilon}_2\right), \tag{7.37}$$

so that

$$\vec{\epsilon}_1\cdot\vec{x}_{nj} = x_{nj}\sin\theta\cos\phi, \quad \vec{\epsilon}_2\cdot\vec{x}_{nj} = x_{nj}\sin\theta\sin\phi, \tag{7.38}$$

and hence

$$\sum_\sigma \left|\langle E_n|\vec{\epsilon}_\sigma(\vec{k})\cdot\hat{\vec{x}}|E_j\rangle\right|^2 = x_{nj}^2\sin^2\theta, \tag{7.39}$$

where θ is the angle between $\vec{k}$ and $\vec{x}_{nj}$. To sum over all directions for $\vec{k}$ we can temporarily view our z axis as being along $\vec{x}_{nj}$. Then θ is the polar angle of $\vec{k}$ and the solid angle is $d\Omega = -d\phi d(\cos\theta)$. The integration over photon propagation directions in the matrix element is x_{nj}^2 multiplied by

$$\int_0^{2\pi} d\phi \int_{-1}^{1} d(\cos\theta)\sin^2\theta = \frac{8\pi}{3}. \tag{7.40}$$

Putting together (7.34), (7.39), and (7.40), the transition rate is given by[7]

$$w = \frac{4\alpha}{3c^2}\omega^3 x_{nj}^2, \quad \omega = \frac{|E_n - E_j|}{\hbar}. \tag{7.41}$$

[6] Our discussion will be imperfect because one really needs a relativistic treatment to properly treat the photon's spin.

[7] This is consistent with [3] and corrects (23.54) in [2].

To summarize, a principal effect of coupling a hydrogenic electron (in an excited atomic state) to the photon field via (7.15) is to cause a transition to a lower energy atomic state with the emission of a photon. The energy and angular momentum of the emitted photon accounts for the difference between the initial and final electron energies and angular momenta. The transition rate is given by (7.41). Of course this is the standard description of spontaneous emission that you have heard of. It is perhaps enlightening to review how many approximations and idealizations are needed to get this familiar scenario. This result only appears at first order in perturbation theory. We need the size of the region in which we are working to be sufficiently large (compared to the wavelength of the emitted photon) to get a quasi-continuum of final photon states. We need the transition time to be sufficiently long compared to atomic time scales so we can use Fermi's Golden rule. Finally we need the emitted photons to have a wavelength much bigger than the Bohr radius so we can use the electric dipole approximation. If any of these approximations are violated the standard phenomenology of spontaneous emission will not necessarily apply.

7.4.3 Lifetime of $2P \to 1S$

Let us use (7.41) to calculate the mean lifetime of the $2P$ state of a hydrogenic atom with respect to decay by spontaneous emission of a single photon. The initial state has the atomic electron in one of its first excited states with non-vanishing angular momentum ($n = 2, l = 1, m = 0, \pm 1$) and zero photons. The final state has the atomic electron in the ground state ($n = 1, l = 0, m = 0$) and a photon with energy $\hbar\omega(k) = E_2 - E_1$. We need to calculate the matrix element (7.36) in (7.41) then average over the three possible initial states. I will only do the computation for the initial state in the $m = 0$ case since the other two cases give the same result thanks to rotational symmetry. So, we want the position vector matrix element with respect to the states (7.9) and (7.10). With $\vec{x} = (r\sin\theta\cos\phi, r\sin\theta\sin\phi, r\cos\theta)$ we get

$$\langle 1,0,0|\hat{\vec{x}}|2,1,0\rangle = \int_0^{2\pi} d\phi \int_0^{\pi} d\theta \int_0^{\infty} dr\, r^2 \sin\theta\, \vec{x} \left(\sqrt{\frac{Z^3}{\pi a^3}} e^{-Zr/a}\right)\left(\sqrt{\frac{Z^3}{32\pi a^3}}\frac{Zr}{a}\cos\theta e^{-Zr/2a}\right) \tag{7.42}$$

$$= (0, 0, \frac{128\sqrt{2}}{243}\frac{a}{Z}). \tag{7.43}$$

The other 2 initial states ($m = \pm 1$) give the same result up to a phase, so the average – the transition rate – is given by

$$w = \frac{131072}{177147}\frac{\alpha a^2 \omega^3}{Z^2 c^2} = \frac{256}{6561} Z^4 \alpha^5 \frac{mc^2}{\hbar}. \tag{7.44}$$

This is the number of transitions per unit time. The mean lifetime τ is the reciprocal of this transition rate, $\tau = 1/w$. With

$$\hbar = 6.582 \times 10^{-16} \text{ eV-s}, \quad mc^2 = 5.110 \times 10^5 \text{ eV}, \quad \alpha = 0.0073 \tag{7.45}$$

we get

$$w = 6.280 \times 10^8 Z^4 \text{ s}^{-1}, \quad \tau = \frac{1.592 \times 10^{-9}}{Z^4} \text{ s}. \tag{7.46}$$

The experimental values I found for the lifetime in the case of hydrogen ($Z = 1$) were

$$\tau = (1.600 \pm 0.004) \times 10^{-9} \text{ s}, \quad \text{(see reference [9])}$$
$$\tau = (1.60 \pm 0.01) \times 10^{-9} \text{ s}, \quad \text{(see reference [10])}$$

These measurements agree well with the theoretical value (7.46), particularly when you consider that we have ignored all kinds of important adjustments to the structure of the atom, for example, spin-orbit coupling (fine structure) and coupling to nuclear spin (hyperfine structure).

7.5 Another perspective

We have seen that the coupling of an atomic electron to the quantum electromagnetic field leads to transitions between the (former) stationary states of the system. This point of view is the one afforded by Dirac's time-dependent perturbation theory which approximates the time dependence of the solutions of the Schrödinger equation in terms of the unperturbed (non-interacting) system. It is interesting to consider the perspective provided by time-*independent* perturbation theory. Recall that this is a way to approximate the spectrum of a (time-independent) Hamiltonian in terms of its unperturbed (non-interacting) form.[8] As mentioned earlier, the interaction between the electron and photons renders the original non-interacting stationary states no longer stationary: the interaction changes the Hamiltonian with a concomitant change in the energy eigenvectors. As far as I know, despite the relative simplicity of the model, there is no known formula for the spectrum of the Hamiltonian (7.15). However, we can consider the approximation of the spectrum via time-independent perturbation theory.

Let me remind you of the salient results of time-independent perturbation theory. If the Hamiltonian $\hat{H}$ takes the form

$$\hat{H} = \hat{H}_0 + \hat{V}, \tag{7.47}$$

with a suitably small perturbation $\hat{V}$, the eigenvalues E_n and eigenvectors $|E_n\rangle$ of $\hat{H}$ can be systematically approximated in terms of those for $\hat{H}_0$, which are denoted by E_n^0, $|E_n^0\rangle$. To first order in the perturbation we have (for non-degenerate eigenvalue E_n^0)

$$E_n \approx E_n^0 + \langle E_n^0|\hat{V}|E_n^0\rangle, \tag{7.48}$$

$$|E_n\rangle \approx |E_n^0\rangle + \sum_{k\neq n} \frac{\langle E_k^0|\hat{V}|E_n^0\rangle}{E_n^0 - E_k^0}|E_k^0\rangle, \tag{7.49}$$

where the sum over k is a sum over a basis of unperturbed energy eigenvectors excluding the one with (non-degenerate) eigenvalue E_n^0. (The approximate energy eigenvector has not been normalized.) There is a bit more to the story, but this will serve for the present discussion.

Let us imagine applying (7.48) and (7.49) to obtain a more accurate description of the ground state of hydrogen. Recall that the unperturbed stationary states are given by (7.11), which are states where the electron is in one of the usual hydrogenic energy levels with various numbers of photons in various stationary states. In particular, the unperturbed ground state is the usual ground state of hydrogen with zero photons: $|\psi_{100}\rangle \otimes |0\rangle$ and the unperturbed ground state energy is -13.6

[8] The relation between these two approximation schemes can be obtained via (2.104) in the case where the Hamiltonian is time independent.

eV. What becomes of this ground state when the interaction with photons is accounted for? From (7.49), the ground state is "dressed" with contributions from the (unperturbed) excited atomic states and various photon states. The contribution of these "virtual" states to the "dressing" of the ground state is (to first order) determined by the matrix element of $\hat{V}$ between the ground state and the other unperturbed stationary states. For example, much as we saw earlier, the matrix element of the $\vec{A} \cdot \vec{p}$ term in the Hamiltonian will, in the electric dipole approximation, dress the ground state with virtual excited hydrogenic states (where $l = 1$ and $m = \pm 1, 0$) and virtual 1-photon states (with any energy-momentum and polarization). The A^2 term in the Hamiltonian will dress the ground state with virtual excited hydrogenic states containing two virtual photons.[9] So, the probability distributions for observables like the position and momentum of the electron will receive small corrections which render them much more intricate than in the non-interacting theory. It is easy to see that the first-order correction to the ground state energy itself vanishes, although it will receive (very small) corrections at the next order in perturbation theory.

One often uses a very picturesque language to describe the state of affairs sketched above. One says the true ground state of the atom involves the atomic electron making "virtual transitions" to and from the unperturbed ground state, emitting and absorbing a cloud of "virtual photons". Of course, one cannot take this language too literally: the virtual processes do not conserve energy and are, in any case, unobservable as real dynamical processes (in contrast to, *e.g.*, spontaneous emission of a real photon). Still, this way of speaking does give some life to the formula (7.49). You can read more about all this in reference [11].

7.6 Problems

1. Consider a "2-state system", that is, a system with a two-dimensional Hilbert space. Suppose the Hamiltonian is

$$\hat{H} = \hat{H}_0 + \hat{V}, \quad \hat{H}_0 = \begin{pmatrix} E_1 & 0 \\ 0 & E_2 \end{pmatrix}, \quad V = \begin{pmatrix} 0 & \gamma \\ \gamma & 0, \end{pmatrix} \tag{7.50}$$

   where E_1, E_2, and γ are some given real numbers. Solve the Schrödinger equation assuming the initial state is the eigenstate of H_0 with eigenvalue E_1. Find the probability that the system is in the eigenstate of H_0 with eigenvalue E_2 at time t. Show how the formula (7.17) defines an approximation to this transition probability.

2. Show that the term quadratic in $\hat{\vec{A}}$ in $\hat{\mathbf{H}}$ is responsible for atomic transitions involving emission or absorption of 2 photons.

3. An atom in the presence of electromagnetic radiation can emit photons by *stimulated emission*. Consider a transition from an atomic state $|E_j\rangle$ in the presence of N photons with polarization σ and wave vector $\vec{k}$ to a state $|E_n\rangle$ in the presence of $N+1$ photons with polarization σ and wave vector $\vec{k}$. Calculate the analog of the matrix element (7.24). Show that for large N this same matrix element can be obtained by applying time dependent perturbation theory just to the hydrogenic atom in the presence of a suitable *classical* electromagnetic field.

4. Show that a hydrogenic transition $2S \to 1S$ by spontaneous emission of a single photon is forbidden to first-order in perturbation theory.

[9] The $\vec{A} \cdot \vec{p}$ term will also contribute two photon states at second order in the perturbation.

5. Calculate the transition rates for $|2,1,\pm 1\rangle \rightarrow |1,0,0\rangle$ and show that they give the same results as we calculated for $|2,1,0\rangle \rightarrow |1,0,0\rangle$.

6. The *parity* of a quantum state of a particle refers to its behavior under the transformation $\vec{x} \rightarrow -\vec{x}$. If a state is unchanged by this transformation it is said to have *even parity*. If it changes sign under this transformation it is said to have *odd parity*. Show that electric dipole transitions necessarily change the parity (even $\leftrightarrow$ odd) of the unperturbed hydrogenic stationary state. Show that $|1,0,0\rangle$ has even parity and $|2,1,0\rangle$ has odd parity.

Chapter 8

Epilogue

The scope of this text has been intentionally rather modest. Indeed, most of the preceding results were already in place by the 1930's. Needless to say, a lot has happened to quantum field theory since then. Let me very briefly mention some of the salient results.[1]

The principal conclusion of decades of experimental and theoretical work in the 20^{th} century was that *everything* can be described by quantum fields. Matter is described in terms of fermionic fields while the (non-gravitational) interactions between matter fields are described by "non-Abelian gauge fields", which are bosonic fields incorporating and generalizing the quantum electromagnetic field. The proper framework for doing all this is the "standard model" of particle physics. The standard model builds all of the known matter from quark fields and lepton fields. Quantum gauge fields represent the strong and electroweak interactions amongst the matter. Finally the standard model includes a spin-0 (scalar) field, known as the *Higgs field*, which allows for particle excitations of various fields to acquire mass in appropriate "low energy" limits.

The majority of predictions of the standard model are obtained via perturbation theory and are thus phrased in terms of interactions between various particles. But some of the more profound physical results are best thought of in terms of the quantum fields themselves. I have in mind, for example, the emergence of the renormalization group as a way of understanding and relating physics at disparate length scales. It is perhaps also worth mentioning that some observed physical effects, *e.g.,* the *Casimir effect*, are normally understood purely in terms of quantum fields – this effect occurs in the vacuum state – and the particle point of view is simply not relevant.

The gravitational interaction is a good deal more interesting from the point of view of quantum field theory. Classically, gravitation is described in terms of the curvature of spacetime via Einstein's general theory of relativity. There are two levels of sophistication one can use when incorporating gravitational effects in quantum field theory (or, if you prefer, when incorporating quantum effects in gravitational field theory). The simplest approach is to try to define quantum field theory in a curved spacetime. This allows one to incorporate the influence of, say, the cosmological expansion, or the event horizon of a black hole, on particle physics. The constructions used in this text tacitly took advantage of the properties of *flat* spacetime. Some of these constructions generalize to a curved spacetime and some do not. Consequently some qualitatively new features arise when incorporating gravitation in this way One of the most striking results of extending quantum field theory to curved

[1] I don't pretend to give a comprehensive – or even objective – summary; the choice of ideas to emphasize and my characterization of them is based upon my personal prejudices.

spacetimes is that the notion of "particles" becomes ambiguous. Depending upon the gravitational field, different observers need not agree on the particle content of the state of a quantum field. Physical manifestations of this phenomenon are particle production by an expanding universe and "Hawking radiation" from black holes. Quantum field theory in curved spacetime underscores the primacy of the notion of quantum field over the notion of particle [12]

Quantum field theory in curved spacetime is designed to treat gravitation classically and everything else quantum mechanically. There are a number of reasons to believe that this framework cannot provide the ultimate description of gravitation. For example, there exist several arguments suggesting that to couple a classical system to a quantum system in this way is at best a limited approximation – and ultimately inconsistent. A more ambitious attempt to incorporate gravitation into the framework of quantum field theory is then to view gravitation itself as a quantum phenomenon.

Starting from Einstein's theory approximated near flat spacetime, it is possible to construct a quantum theory of a non-interacting gravitational field, quite analogous to what we did with the electromagnetic field. This quantum gravitational field has massless helicity-2 particle excitations, usually called "gravitons". In perturbation theory one can build the interaction of gravitons with particles and recover (perturbatively) the classical gravitational interaction of Einstein's theory. Challenges occur when one tries to go beyond this point and incorporate genuine quantum effects. Technically, the perturbation expansion only defines the quantum gravity theory in conjunction with measurement of an infinite number of coupling constants – the theory is "non-renormalizable". Alternatively, one still needs to find a high energy/short distance quantum field theory to "complete" the quantum field theory which is approximated by Einstein gravity at sufficiently large distance scales. Conceptual problems also arise from the fact that the spacetime geometry in a putative quantum field theory of gravity should somehow be undergoing quantum fluctuations. Normally, the (classical) spacetime geometry is used to build and interpret a quantum field theory, so it becomes problematic to construct a quantum field theory of the spacetime geometry itself. The challenges inherent in devising a quantum description of the gravitational interaction represent, in many ways, the principal problem of theoretical physics. Without some compelling guidance from experiment, it is likely to remain an open problem for some time to come.

Bibliography

[1] P. A. M. Dirac, *The Principles of Quantum Mechanics*, Oxford Science Publications (1958).

[2] E. Merzbacher, *Quantum Mechanics*, John Wiley & Sons (1998).

[3] J. Sakurai, *Advanced Quantum Mechanics*, Addison-Wesley (1982).

[4] A. Einstein, B. Podolsky, N. Rosen, "Can Quantum-Mechanical Description of Physical Reality be Considered Complete?" *Phys. Rev.* **47**, 777–780 (1935).

[5] C. G. Torre, *Introduction to Classical Field Theory*, https://digitalcommons.usu.edu/lib_mono/3/ (2017).

[6] T. Newton and E. Wigner, "Localized States for Elementary Systems", *Rev. Mod. Phys.* **21**, 845 (1949).

[7] A. S. Wightman, "On the Localizability of Quantum Mechanical Systems", *Rev. Mod. Phys.* **34**, 845 (1962).

[8] T. Jordan, *J. Math. Phys* **19**, 1382 (1978).

[9] W. Bickel and A. Goodman, *Phys. Rev.* **148**, 1 (1966).

[10] E. Chupp, L. Dotchin, and D. Pegg, *Phys. Rev.* **175** 44 (1968).

[11] R. Passante and E. Power, *Phys. Rev. A* **35**, 188 (1987).

[12] P. C. W. Davies, "Particles Do Not Exist," in *Quantum Theory Of Gravity*, 66-77 (Adam Hilger, 1984).