Quantum Optics for the Impatient

Morgan W. Mitchell
ICFO - Institut de Ciències Fotòniques
Castelldefels (Barcelona)

Contents

Preface

This text began as notes for the course "Experimental Quantum Optics and Quantum Information," attended by students from ICFO and the Barcelona-area universities UAB, UB, and UPC. When the course began, several comprehensive and high-quality books had recently been published on quantum optics. These books present, in a complete and coherent fashion, results from decades of work in quantum optics before quantum information became important. They might be compared to Max Born and Emil Wolf's "Principles of Optics," which describes the state of knowledge in optics before the invention of the laser. These books should not be overlooked. Any serious student of quantum optics must be familiar with at least one such authoritative text.

Then why write a new text on Quantum Optics?

Recent progress in quantum optics (QO) has largely been related to quantum information (QI) and quantum technologies (QT): communications and information processing based on the unique features of quantum mechanics. The experimental techniques of quantum optics, which include the precise generation, manipulation, and measurement of quantum states of light, are very well suited to experiments in quantum information. Many tasks in quantum information were first demonstrated with optical implementations. The theory of quantum optics, however, can seem pretty foreign to a practitioner of QI, because QO comes from quantum field theory while QI is from ordinary quantum mechanics. Thus, many students (and others new to the field) arrive with an interest to understand quantum optics, not for itself, but as a tool for doing (or understanding) experimental quantum information. Typically these people are in a hurry. Thus the need for a rapid introduction, a "quick-start" manual, for the area of quantum optics.

These notes aim to provide a self-contained introduction to quantum optics, for a readership that is comfortable with quantum mechanics, electromagnetism, modern optics, and the associated mathematics. The text presents the core elements of quantum optics theory, the ones most likely to be encountered in experimental work or in related theory, in a manner that aims to build physical intuition, and will be useful for simple calculations. The text does *not* aim to be comprehensive. Rather, we hope the reader will look to the aforementioned reference texts for extensive discussions, historical references, and authoritative formulations. We hope these texts will be both more interesting and more accessible after this introduction.

Recommended Background

Many fields have contributed to the development of quantum optics, and some prior understanding of these fields is necessary to fully appreciate what happens in quantum optics experiments. Most important are physical optics, nonlinear optics, quantum mechanics, and some basic notions from quantum field theory. Also important are signal theory, atomic physics, electronics, detector physics, laser theory. The reader is strongly recommended to consult the books listed below when background information is needed.

Physical Optics and Optical Technologies

Fundamentals of Photonics by B. E. A. Saleh and M. C. Teich, Wiley, 1991.

Nonlinear Optics

Nonlinear Optics, 2nd Ed. by R. W. Boyd, Academic Press, 2002.

Atomic Physics

Laser Spectroscopy: Basic Concepts and Instrumentation by W. Demtröder, Springer, 2000.

Atomic Physics by C. J. Foot, Oxford, 2005.

Laser Physics

Quantum Electronics, 3rd Ed by A. Yariv, Wiley, 1989.

Lasers by A. E. Siegman, University Science Books, 1986.

Laser Physics, New Ed. by M. Sargent, M. O. Scully, W. E. Lamb, Perseus, 1974.

Lasers, 4th Ed. by O. Svelto, Springer, 2004.

Lasers, by P. W. Milonni and J. H. Eberly, Wiley, 1988.

Advanced Mechanics

Classical Mechanics, 3rd Ed. by H. Goldstein, C. P. Poole and J. L. Safko Addison-Wesley, 2002.

Electrodynamics

Classical Electrodynamics, 3rd Ed. by J. D. Jackson, Wiley, 1998.

Quantum Mechanics

Modern Quantum Mechanics by J. J. Sakurai, Addison-Wesley, 1985.

Advanced Quantum Mechanics by J. J. Sakurai, Addison-Wesley, 1967.

The Quantum Vacuum: An Introduction to Quantum Electrodynamics by P. W. Milonni, Academic Press, 1993.

Quantum Optics Textbooks

The books listed below are interesting either because they are historical and authoritative, or because they are recent and up-to-date. Almost all cover quantum optics in more detail than this text. The reader is strongly encouraged to follow at least one of these books *at the same time* as reading this text. Much of what is contained in these notes is intended to illustrate or explain what is contained, in denser form, in the books.

Quantum Optics by M. O. Scully and M. S. Zubairy , Cambridge, 1997.

Quantum Optics by D. F. Walls and G. J. Milburn, Springer, 1995.

A Guide to Experiments in Quantum Optics, 2nd Ed. by H-A. Bachor and T. C. Ralph, Wiley, 2004.

The Quantum Theory of Light, 3rd Ed. by R. Loudon, Oxford, 2000.

Optical Coherence and Quantum Optics by L. Mandel and E. Wolf, Cambridge, 1995.

Elements of Quantum Optics, 3rd Ed. by P. Meystre, M. Sargent, Springer, 2006.

Quantum Optics in Phase Space by W. P. Schleich, Wiley, 2001.

Quantum Optics, An Introduction by M. Fox, Oxford, 2006.

Introductory Quantum Optics by C. Gerry, and P. L. Knight Cambridge, 2004.

Methods in Theoretical Quantum Optics by S. M. Barnett and P. M. Radmore, Oxford, 2003.

Quantum Optics by W. Vogel and D-G Welsch, Wiley, 2006.

Fundamentals of Quantum Optics and Quantum Information by P. Lambropoulos, D. Petrosyan, Springer, 2006.

Quantum Optics by J. C. Garrison and R. Y. Chiao, Oxford, 2008

Chapter 1

Introduction

1.1 What is quantum optics?

Probably the simplest workable definition for the term "quantum optics" is "the part of optics that requires quantum physics." With this definition, we could say that splitting sunlight into its component colours with a prism is not quantum optics, because classical electromagnetic theory is enough to explain this phenomenon. But many optical phenomena do require quantum physics for their explanation. Historically important examples are the blackbody radiation spectrum, the photoelectric effect, and Compton scattering. Quantum physics really became central to optics with the invention of the MASER and LASER in the 1950s and 1960s. These instruments are inherently quantum mechanical - they operate based on stimulated emission of photons from material with quantized energy levels. Lasers, which emit powerful, coherent radiation, also make it much easer to see quantum effects in materials, for example Rabi oscillation between different energy levels. Modern photodetectors based on electron-hole pair creation in semiconductors are also clearly quantum mechanical.

Another definition, more relevant to this text, is "the part of optics that requires a quantum description of optical fields." This more restrictive definition would still include Compton scattering and the blackbody spectrum, but would exclude effects like Rabi oscillation and photodetection, which can be explained with a semi-classical description. In semi-classical optics, the electromagnetic field is described classically and the material systems are described quantum mechanically. Interestingly, the photo-electric effect can be explained semi-classically and would not be "quantum optics" under this definition. Historical examples of phenomena requiring a quantum description of the fields include the Lamb shift, spontaneous emission of photons, laser operation, and many phenomena involving the statistical behaviour of optical particles (photons) and optical fields.

This last category of effects is worth describing in more detail. A paradigmatic example is photon anti-bunching. "Bunching" describes the tendency of things to cluster together. Concerning light, we can imagine a light source that emits photons in groups, e.g. as pulses of light. This would be "photon bunching." We can also imagine a source that emits photons one at a time,

so that two photons are unlikely to arrive together. This is "photon anti-bunching." If we try to describe these situations with the semi-classical theory, we run into a problem. The semi-classical theory describes the interaction of a classical field with quantized matter, for example in the detectors. A consequence of this is that if one electron in a detector can be excited by the field, its neighbors will experience the same field, and they also can be excited. So whenever there is the possibility for one detection event, there will be the possibility for more detection events at the same time. This difference between the semi-classical and fully-quantum descriptions of light was first put to the test in 1977, when photon anti-bunching was first observed.

The generation of statistical behaviours that cannot be produced by classical physics is one of the main reasons for the current interest in quantum optics. This is closely related to quantum technologies, which (as a broad generalization) aim to use uniquely quantum behaviours in applications involving information. One well-known example is the use of single photons to generate secret keys shared between people in distant locations. This benefits from the anti-bunching just described, because if the photon reaches its intended recipient, it can't also have been picked up by an eavesdropper. Another example is the manipulation of photon correlations to evade sensing limits like the shot noise limit. "Shot noise" describes the behaviour of uncorrelated particles, and is a prediction of the semi-classical theory. Quantum optics does not need to obey this limit. The quantum optical effect of "squeezed light" has recently been used to improve the sensitivity of gravitational wave detectors, which are giant optical interferometers.

Chapter 2

Foundations

We begin with the quantization of the electromagnetic field. "Quantization" in this context means inventing a quantum theory that reproduces the results of classical electromagnetism in the classical limit. I say "inventing" rather than "deriving" because in fact there is no deterministic way to turn a classical theory into a correct quantum theory. Nevertheless, we will see that the choice is natural, and there is little question that we have the right theory.

The procedure that we use is called "canonical quantization," and proceeds from the equations of motion for light (Maxwell's equations), to a Lagrangian, to an operator representation of the fields. Before we quantize the electromagnetic field, we first quantize something simpler, the harmonic oscillator. In fact, we will see that the electromagnetic field *is* in effect a collection of harmonic oscillators, so the results will be useful immediately.

2.1 Simple harmonic oscillator

The classical simple harmonic oscillator obeys the following second-order ordinary differential equation

$$\ddot{x} = -\omega^2 x \tag{2.1}$$

where x is the position and ω is the angular frequency of oscillation. This equation can be derived from the Lagrangian

$$L = \frac{m}{2}\dot{x}^2 - \frac{m\omega^2}{2}x^2 \tag{2.2}$$

by applying the Euler-Lagrange equation

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{x}} = \frac{\partial L}{\partial x} \tag{2.3}$$

Here m is a constant which turns out to be the mass.

The canonical momentum conjugate to x is

$$p_x \equiv \frac{\partial L}{\partial \dot{x}} = m\dot{x}. \tag{2.4}$$

The Hamiltonian is

$$H \equiv \sum_i p_i \dot{q}_i - L = \frac{p^2}{2m} + \frac{m\omega^2}{2}x^2. \tag{2.5}$$

Note that in this quantization procedure, the equations of motion are fundamental, not the Lagrangian or Hamiltonian. Classical theories such as Newton's laws, Maxwell's equations, or fluid dynamics, are based in equations of motion. The Lagrangian and Hamiltonian are secondary, chosen to give the equations of motion.

To create the quantum theory of the harmonic oscillator, we keep this Hamiltonian operator and we identify x and p_x as observables and associate them with the operators $\hat{x}$ and $\hat{p}_x$. In this way the classical Hamiltonian becomes the Hamiltonian operator

$$\hat{H} = \frac{m\omega^2}{2}\hat{x}^2 + \frac{1}{2m}\hat{p}_x^2. \tag{2.6}$$

Finally, we assume that these operators have the commutation relation $[\hat{x}, \hat{p}_x] = i\hbar$. This implies an uncertainty relation $\delta x \delta p_x \geq \hbar/2$. This is the heart of the canonical quantization procedure: we assume that canonically conjugate coordinates and momenta have the commutation relation $[q, p_q] = i\hbar$, which replaces the classical relationship involving the Poisson bracket $\{q, p_q\}_{\text{PB}} = 1$.

We note that we can calculate the equations of motion for x and p two ways (and get the same result). Classically, the Hamilton-Jacobi equations of motion

$$\dot{q} = \frac{\partial H}{\partial p}, \quad \dot{p} = -\frac{\partial H}{\partial q} \tag{2.7}$$

give

$$\begin{aligned} \dot{x} &= \frac{1}{m}p \\ \dot{p} &= -m\omega^2 x \end{aligned} \tag{2.8}$$

Quantum mechanically, the Heisenberg equation of motion

$$\dot{A} = \frac{1}{i\hbar}[A, H] \tag{2.9}$$

(valid for any operator A that does not explicitly depend on time) gives

$$\begin{aligned} \dot{x} &= \frac{1}{2im\hbar}[x, p^2] = \frac{1}{m}p \\ \dot{p} &= \frac{m\omega^2}{2i\hbar}[p, x^2] = -m\omega^2 x. \end{aligned} \tag{2.10}$$

More generally, if we have a multi-dimensional system with several coordinates q_i and their conjugate momenta p_{q_i}, then we assume the commutation relations $[q_i, q_j] = [p_{q_i}, p_{q_j}] = 0$ and $[q_i, p_{q_j}] = i\hbar\delta_{ij}$ where δ_{ij} is the Kronecker delta. This implies there is an uncertainty relationship only between canonically conjugate variables.

2.2 Quantization of the electromagnetic field

2.2.1 Classical equations of motion

We start with a description of light in empty space, either vacuum or the (empty) inside of an optical resonator defined by reflecting surfaces such as mirrors. The equations of motion are the source-free Maxwell equations

$$\begin{aligned} \nabla \cdot \mathbf{E} &= 0 && (2.11)\\ \nabla \cdot \mathbf{B} &= 0 && (2.12)\\ \nabla \times \mathbf{E} &= -\frac{\partial \mathbf{B}}{\partial t} && (2.13)\\ \nabla \times \mathbf{B} &= \mu_0 \varepsilon_0 \frac{\partial \mathbf{E}}{\partial t} && (2.14) \end{aligned}$$

These are simpler in terms of the vector potential $\mathbf{A}$ (taken in the Coulomb gauge $\nabla \cdot \mathbf{A} = 0$) which satisfies

$$\begin{aligned} \mathbf{B} &= \nabla \times \mathbf{A}\\ \mathbf{E} &= -\frac{\partial \mathbf{A}}{\partial t} && (2.15) \end{aligned}$$

Substituting into 2.14, we find the wave equation for $\mathbf{A}$

$$\left(\nabla^2 - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\right)\mathbf{A} = \mathbf{0} \qquad (2.16)$$

It is convenient at this point to expand the spatial part of the vector potential in vector spatial modes $\mathbf{u}_{k,\alpha}$ defined by

$$\nabla^2 \mathbf{u}_{k,\alpha}(\mathbf{r}) = -k^2 \mathbf{u}_{k,\alpha}(\mathbf{r}) \qquad (2.17)$$

where k is the wave-number and $\alpha = 1, 2$ is an index for the polarization. If we choose these modes well, they will be orthonormal, $\int \mathrm{d}^3 r\, \mathbf{u}^*_{k,\alpha}(\mathbf{r}) \cdot \mathbf{u}_{k',\alpha'}(\mathbf{r}) = \delta_{k,k'}\delta_{\alpha,\alpha'}$. Thus the vector potential is

$$\mathbf{A}(\mathbf{r}, t) = \sum_{k,\alpha} q_{k,\alpha}(t)\mathbf{u}_{k,\alpha}(\mathbf{r}) \qquad (2.18)$$

where the $q_{k,\alpha}$ are time-varying mode amplitudes. Substituting into equation (2.16), we find

$$\ddot{q}_{k,\alpha} = -c^2 k^2 q_{k,\alpha} \equiv -\omega_k^2 q_{k,\alpha} \qquad (2.19)$$

which is precisely the same form as equation (2.1). Because the equations of motion are the same, we use the same Lagrangian, and arrive at the same canonical momentum and Hamiltonian. The momentum is $p_{k,\alpha} = m\dot{q}_{k,\alpha}$. The single-mode Hamiltonian is

$$H_{k,\alpha} = \frac{1}{2}m\omega^2 q_{k,\alpha}^2 + \frac{1}{2m}p_{k,\alpha}^2. \tag{2.20}$$

The "mass" m in this equation needs a little explanation. It is not present in the equations of motion, so it is not determined by the classical dynamics. It is in fact a parameter we are free to choose. As we will see, the right choice for the "mass" is $m = \varepsilon_0$, where ε_0 is the permittivity of free space. We also note that the electric field is

$$\mathbf{E}(\mathbf{r},t) = -\frac{\partial}{\partial t}\mathbf{A}(\mathbf{r},t) = -\sum_{k,\alpha}\dot{q}_{k,\alpha}(t)\mathbf{u}_{k,\alpha}(\mathbf{r}) = -\frac{1}{\varepsilon_0}\sum_{k,\alpha}p_{k,\alpha}(t)\mathbf{u}_{k,\alpha}(\mathbf{r}). \tag{2.21}$$

Thus for each mode $\mathbf{u}_{k,\alpha}$, the vector potential amplitude $x_A \equiv q_{k,\alpha}$ is canonically conjugate to $-\varepsilon_0 x_E$ where $x_E \equiv p_{k,\alpha}/\varepsilon_0$ is the electric field amplitude. We now quantize the theory by replacing the c-numbers $q_{k,\alpha}, p_{k,\alpha}$ with operators $\hat{q}_{k,\alpha}, \hat{p}_{k,\alpha}$ which obey the commutation relation $[\hat{q}_{k,\alpha}, \hat{p}_{k,\alpha}] = i\hbar$. This immediately implies an uncertainty relation for each mode of the **A** and **E** fields $\delta x_A \delta x_E \geq \hbar/2\varepsilon_0$.

As we have said, each mode of the field *is* a harmonic oscillator: it has the same classical dynamics and the same quantum theory. We remind ourselves of some results from the theory of harmonic oscillators. We work in the Heisenberg representation so that the operators evolve according to the Heisenberg equation of motion $d\hat{A}/dt = (1/i\hbar)[\hat{A}, \hat{H}]$.

Hamiltonian	$\hat{H} = \frac{1}{2}m\omega^2\hat{x}^2 + \frac{1}{2m}\hat{p}^2 = \hbar\omega(\hat{n} + 1/2)$
Number states	$\|n=0\rangle, \|n=1\rangle, \|n=2\rangle, ...$
Annihilation operator	$\underline{\hat{a}}(t) = \hat{a}\exp[-i\omega t]$
	$\hat{a}\|n\rangle = \sqrt{n}\|n-1\rangle$
Creation operator	$\underline{\hat{a}}^\dagger(t) = \hat{a}^\dagger\exp[i\omega t]$
	$\hat{a}^\dagger\|n\rangle = \sqrt{n+1}\|n+1\rangle$
Number operator	$\hat{n} = \hat{a}^\dagger\hat{a}$
position operator	$\hat{x}(t) = \sqrt{\frac{\hbar}{2m\omega}}(\hat{a}e^{-i\omega t} + \hat{a}^\dagger e^{i\omega t})$
momentum operator	$\hat{p}(t) = -i\sqrt{\frac{\hbar\omega m}{2}}(\hat{a}e^{-i\omega t} - \hat{a}^\dagger e^{i\omega t})$
commutation relations	$[\hat{x}(0), \hat{p}(0)] = i\hbar$
	$[\hat{a}, \hat{a}^\dagger] = 1$

Summary of harmonic oscillator states and operators. Note that we have used the underlined symbols $\underline{\hat{a}}$ and $\underline{\hat{a}}^\dagger$ to indicate the time-varying Heisenberg-picture operators, and we use the ordinary symbols $\hat{a} \equiv \underline{\hat{a}}(t=0)$ and $\hat{a}^\dagger \equiv \underline{\hat{a}}^\dagger(t=0)$ to indicate the static operators. For example, $\hat{a}\|n\rangle = \sqrt{n}\|n-1\rangle$ while $\underline{\hat{a}}\|n\rangle = \exp[-i\omega t]\sqrt{n}\|n-1\rangle$.

Finally, we express the quantized vector potential in terms of creation and annihilation operators

$$\hat{\mathbf{A}}(\mathbf{r},t) = \sum_{k,\alpha}\sqrt{\frac{\hbar}{2\omega_k\varepsilon_0}}\left(\hat{a}_{k,\alpha}\mathbf{u}_{k,\alpha}(\mathbf{r})e^{-i\omega_k t} + \hat{a}_{k,\alpha}^\dagger\mathbf{u}_{k,\alpha}^*(\mathbf{r})e^{i\omega_k t}\right). \tag{2.22}$$

The quantized electric field $\hat{\mathbf{E}} = -\partial\hat{\mathbf{A}}/\partial t$ is

$$\hat{\mathbf{E}}(\mathbf{r}, t) = i\sum_{k,\alpha}\sqrt{\frac{\hbar\omega_k}{2\varepsilon_0}}\left(\hat{a}_{k,\alpha}\mathbf{u}_{k,\alpha}(\mathbf{r})e^{-i\omega_k t} - \hat{a}^\dagger_{k,\alpha}\mathbf{u}^*_{k,\alpha}(\mathbf{r})e^{i\omega_k t}\right) \tag{2.23}$$

In the case that we are dealing with fields in free space (no resonator to define the modes $\mathbf{u}$), it is conventional to define a "box" of volume L^3 to define the modes $\mathbf{u}_{k,\alpha}(\mathbf{r}) = \mathbf{e}_\alpha \exp[i\mathbf{k}\cdot\mathbf{r}]/\sqrt{L^3}$ where $\mathbf{e}_\alpha$ are polarization vectors perpendicular to $\mathbf{k}$. In this case the fields are

$$\hat{\mathbf{A}}(\mathbf{r}, t) = \sum_{k,\alpha}\sqrt{\frac{\hbar}{2\omega_k\varepsilon_0 L^3}}\left(\mathbf{e}_\alpha\hat{a}_{k,\alpha}e^{i\mathbf{k}\cdot\mathbf{r}}e^{-i\omega_k t} + \mathbf{e}^*_\alpha\hat{a}^\dagger_{k,\alpha}e^{-i\mathbf{k}\cdot\mathbf{r}}e^{i\omega_k t}\right) \tag{2.24}$$

and

$$\hat{\mathbf{E}}(\mathbf{r}, t) = i\sum_{k,\alpha}\sqrt{\frac{\hbar\omega_k}{2\varepsilon_0 L^3}}\left(\mathbf{e}_\alpha\hat{a}_{k,\alpha}e^{i\mathbf{k}\cdot\mathbf{r}}e^{-i\omega_k t} - \mathbf{e}^*_\alpha\hat{a}^\dagger_{k,\alpha}e^{-i\mathbf{k}\cdot\mathbf{r}}e^{i\omega_k t}\right) \tag{2.25}$$

The quantized magnetic field $\hat{\mathbf{B}} = \nabla\times\hat{\mathbf{A}}$ is then

$$\hat{\mathbf{B}}(\mathbf{r}, t) = i\sum_{k,\alpha}\sqrt{\frac{\mu_0\hbar\omega_k}{2L^3}}\left(\mathbf{f}_\alpha\hat{a}_{k,\alpha}e^{i\mathbf{k}\cdot\mathbf{r}}e^{-i\omega_k t} - \mathbf{f}^*_\alpha\hat{a}^\dagger_{k,\alpha}e^{-i\mathbf{k}\cdot\mathbf{r}}e^{i\omega_k t}\right) \tag{2.26}$$

where $\mathbf{f}_\alpha = \mathbf{e}_\alpha\times\mathbf{k}/|\mathbf{k}|$ is the magnetic field polarization vector. Using equations (2.25) and (2.26) it is straightforward to verify that the total Hamiltonian describing each mode as a harmonic oscillator,

$$\hat{H} = \sum_{k,\alpha}\hat{H}_{k,\alpha} = \sum_{k,\alpha}\frac{1}{2}m\omega^2\hat{q}^2_{k,\alpha} + \frac{1}{2m}\hat{p}^2_{k,\alpha} = \sum_{k,\alpha}\hbar\omega_k(\hat{a}^\dagger_{k,\alpha}\hat{a}_{k,\alpha} + \frac{1}{2}) \tag{2.27}$$

agrees with the usual electro-magnetic Hamiltonian

$$\hat{H}_{EM} = \frac{1}{2}\int d^3r\,\left(\varepsilon_0|\hat{E}|^2 + \frac{1}{\mu_0}|\hat{B}|^2\right) = \sum_{k,\alpha}\hbar\omega_k(\hat{a}^\dagger_{k,\alpha}\hat{a}_{k,\alpha} + \frac{1}{2}). \tag{2.28}$$

In fact, this agreement is achieved because we choose $m = \varepsilon_0$, as mentioned above.

2.3 Quadratures

Although the vector potential is more fundamental (at least for quantum field theory), in optics we almost always work with the electric field. This is because most materials interact more strongly with the electric field than with the magnetic field, and because the vector potential is not very "physical" (it is not gauge invariant, for example). We would like to forget about the vector potential, but somehow keep the quantum physics that is summarized in the uncertainty relationship $\delta x_A\delta x_E \geq \hbar/2\varepsilon_0$. Can we describe everything we need in terms of just the field $\mathbf{E}$? In

fact we can: For a harmonic oscillator, the position and momentum are always one quarter cycle out of phase. Because of this, if we describe the amplitude of the electric field now, and also a quarter cycle later, we effectively describe both **E** and **A**. Classically, we would write the electric field in terms of two *quadrature amplitudes* X_1, X_2 as $E(\mathbf{r}, t) = X_1 \sin(\omega t - \mathbf{k}\cdot\mathbf{r}) - X_2 \cos(\omega t - \mathbf{k}\cdot\mathbf{r})$. Here, we define two *quadrature operators* $\hat{X}_1, \hat{X}_2$ through[1]

$$\hat{E}(\mathbf{r}, t) = \sqrt{\frac{\hbar\omega_k}{2\varepsilon_0 L^3}} \left[\hat{X}_1 \sin(\omega t - \mathbf{k}\cdot\mathbf{r}) - \hat{X}_2 \cos(\omega t - \mathbf{k}\cdot\mathbf{r})\right]. \tag{2.29}$$

For this to agree with a single mode's contribution to equation (2.25)

$$\hat{E}(\mathbf{r}, t) = i\sqrt{\frac{\hbar\omega_k}{2\varepsilon_0 L^3}} \left(\hat{a}_{k,\alpha} e^{i\mathbf{k}\cdot\mathbf{r}} e^{-i\omega_k t} - \hat{a}^\dagger_{k,\alpha} e^{-i\mathbf{k}\cdot\mathbf{r}} e^{i\omega_k t}\right) \tag{2.30}$$

we must have

$$\hat{X}_1 = \hat{a} + \hat{a}^\dagger \tag{2.31}$$
$$\hat{X}_2 = i(\hat{a}^\dagger - \hat{a}). \tag{2.32}$$

The quadrature operators are hermitian, and thus observable. In fact, $\hat{X}_1$ is proportional to the vector potential amplitude $\hat{x}_A$ at one instant in time and $\hat{X}_2$ is proportional to electric field amplitude $\hat{x}_E$ at the same instant in time. They have the commutation relation

$$[\hat{X}_1, \hat{X}_2] = 2i \tag{2.33}$$

and uncertainty relation

$$\delta\hat{X}_1 \delta\hat{X}_2 \geq 1. \tag{2.34}$$

Lastly, the Hamiltonian is

$$\hat{H} = \hbar\omega(\hat{a}^\dagger\hat{a} + \frac{1}{2}) = \frac{\hbar\omega}{4}(\hat{X}_1^2 + \hat{X}_2^2). \tag{2.35}$$

At just one point in space (or in fact anywhere along a phase front) $\mathbf{k}\cdot\mathbf{r}$ is a constant. Without loss of generality we choose a point where $\mathbf{k}\cdot\mathbf{r} = 0$. At this point the electric field is

$$\hat{E}(0, t) \propto \hat{X}_1 \sin(\omega t) - \hat{X}_2 \cos(\omega t). \tag{2.36}$$

This would be the field experienced by a stationary atom, for example. The quadratures X_1 and X_2 are simply two coefficients in the Fourier decomposition of the field $E(0, t)$.

[1] From here on, we are just considering one mode, so we leave out the mode indices k, α and the polarization. We are also explicitly considering a traveling wave, because that is the most familiar situation. A very similar derivation can be made assuming standing waves proportional to $X_1 u(\mathbf{r}) \sin(\omega t) - X_2 u(\mathbf{r}) \cos(\omega t)$.

2.4 Connection to classical theory

We have finished with the quantization of the electromagnetic field. The equations above describe the electric and magnetic field operators which are the observables of the quantum theory of light. Each mode of the field is a harmonic oscillator, and for this reason we have expanded the field operators in modes and written them in terms of creation and annihilation operators. We also introduced quadrature operators to express the uncertainty relations entirely in terms of the electric field. Written this way, the theory does not *look* very similar to classical electromagnetism, but in fact the two theories are very similar. For example, the Maxwell equations are still true. They describe the evolution of the field operators (in the Heisenberg representation, of course)

$$\nabla \cdot \hat{\mathbf{E}} = 0 \tag{2.37}$$

$$\nabla \cdot \hat{\mathbf{B}} = 0 \tag{2.38}$$

$$\nabla \times \hat{\mathbf{E}} = -\frac{\partial \hat{\mathbf{B}}}{\partial t} \tag{2.39}$$

$$\nabla \times \hat{\mathbf{B}} = \mu_0 \varepsilon_0 \frac{\partial \hat{\mathbf{E}}}{\partial t}. \tag{2.40}$$

An immediate consequence of this is that the classical values for the field are still correct, in a sense: they are the expectation values for the quantum fields[2]. The quantum theory is different in two key ways. First, the uncertainty principle applies, between the $\hat{\mathbf{A}}$ and $\hat{\mathbf{E}}$ fields or between the $\hat{X}_1$ and $\hat{X}_2$ quadratures, leading to uncertainty and quantum noise. A great deal of work has been done to understand, measure, and manipulate quantum noise, for fundamental understanding of quantum mechanics, but also to make more sensitive measurements. Second, quantum fields can have a rich variety of states: number states, coherent states, squeezed states, entangled states, etc. while the classical theory can only have classical values. It is this variety of states that makes quantum optics interesting for encoding quantum information and we now pass to describing these states.

[2]Note that while the average values of the *fields* are the same in the quantum and classical theories, the averages of other quantities may not be. Consider for example the *intensity* detected at the output of an optical amplifier when no light is injected at the input. Classically, the input field is zero and the output field is zero, which implies zero output intensity also. A real amplifier, however, will output a nonzero intensity, due to amplified spontaneous emission. Quantum mechanically, the input field is the vacuum state, which includes vacuum fluctuations about a zero average value. This is amplified to give detectable light at the output. The average output field is still zero, but the intensity is not.

Chapter 3

Quantum states of light

3.1 Photons

The Hamitonian is $\hat{H} = \hbar\omega(\hat{a}^\dagger\hat{a} + 1/2)$. Based on Planck's hypothesis we believe that a photon has energy $\hbar\omega$, so we interpret the $\hat{n} = \hat{a}^\dagger\hat{a}$ as the number of photons in the mode. This means that $\hat{a}(0)$ destroys a photon, and $\hat{a}^\dagger(0)$ creates one. This is why they're called creation and annihilation operators, after all.

3.2 Vacuum

The ground state of the field is the "vacuum state" $|0\rangle$ defined by $\langle 0|\,\hat{a}^\dagger\hat{a}\,|0\rangle = 0$. It has non-zero energy $E_{\text{vac}} = \hbar\omega/2$ and fluctuations $(\Delta X_{1,2})^2 = <\hat{X}_{1,2}^2> - <\hat{X}_{1,2}>^2 = 1$. Thus it is a minimum uncertainty state $\delta X_1 \delta X_2 = 1$.

3.3 Number states

The number states, or "Fock states" are defined by

$$|n\rangle \equiv \frac{(\hat{a}^\dagger)^n}{\sqrt{n!}}\,|0\rangle \tag{3.1}$$

or $\hat{n}\,|n\rangle = n\,|n\rangle$. These are energy eigenstates with energy $\hbar\omega(n + 1/2)$. The number states are complete and orthonormal, and for many problems, especially those involving photon counting, they are the most natural basis to use. They are, however, very far from classical behaviour. For example, the expectation values of the quadratures are $\langle n|\,\hat{X}_{1,2}\,|n\rangle = 0$, while the variances are $(\Delta X_{1,2})^2 = <\hat{X}_{1,2}^2> - <\hat{X}_{1,2}>^2 = 2n + 1$. Viewed in terms of quadratures, number states consist entirely of noise.

3.4 Coherent states

The energy eigenstates (number states) have zero average field. Clearly this isn't the case when we turn on a laser or a microwave oven. Is there a quantum state that behaves like an oscillating electric field? As in ordinary quantum mechanics, in order for an observable to oscillate, there has to be a superposition of at least two states with different energies. In the case of the electric field (or the quadratures) this means there has to be a superposition of different numbers of photons. What about a state like

$$|\psi\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle), \tag{3.2}$$

does this have an oscillating average field? It is easy to show that $\langle\psi|\,\hat{X}_1\,|\psi\rangle = 1$ and $\langle\psi|\,\hat{X}_2\,|\psi\rangle = 0$ so that

$$\langle\psi|\,\hat{\mathbf{E}}(t)\,|\psi\rangle \propto \sin(\omega t). \tag{3.3}$$

So yes, a superposition of energy eigenstates does oscillate. In fact, any field state that looks at all classical (that has a nonzero expectation value for the E field) must have an indeterminate number of photons.[1]

So what sort of field does a laser (or a radio station for that matter) actually produce? We think that the classical description of the E-M field should be pretty much correct in these cases because there are so many photons involved. We want to find a quantum state that is as classical as possible.

The "most nearly classical" states should have minimum uncertainty $\delta X_1 \delta X_2 = 1$ and should oscillate like the classical field. It turns out that these states are eigenstates of the annihilation operator $\hat{a}\,|\alpha\rangle = \alpha\,|\alpha\rangle$. The name "coherent states" was given to this group of states by Roy Glauber, who first wrote about them in connection with quantum optics.

A coherent state $|\alpha\rangle$ can be expressed in the number basis as

$$|\alpha\rangle = e^{-|\alpha|^2/2} \sum_{n=0}^{\infty} \frac{\alpha^n}{\sqrt{n!}}\,|n\rangle\,. \tag{3.4}$$

[1] This is especially strange when you realize that most particles are not *allowed* to have an indeterminate number (at least you can't get away with hypothesizing the zero/one state above). For example, conservation of lepton number means that while you can lose an electron from the universe, you're guaranteed to create or destroy at least one other particle (of the electron or neutrino sort) in the process. Your state could be $(|0e^- > |1\nu_e > + |1e^- > |0\nu_e >)/\sqrt{2}$, but that's not the same thing.

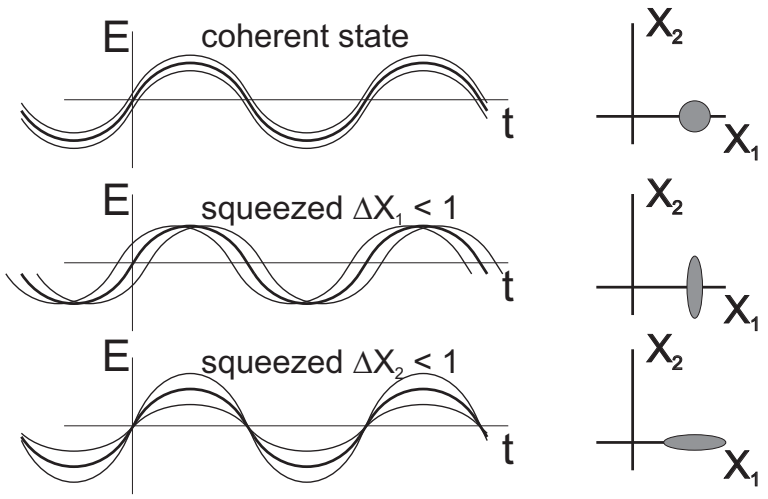

Figure 3.1: Left plots: Average value and uncertainty for a coherent state (top) and two squeezed states with the same average values for the field. In each plot, the heavy line indicates the average field value $< E(t) >$ while the light lines indicate the average plus/minus $\Delta E(t)$. Right plots: uncertainty ellipse representations of these states.

Coherent states have some nice properties.

$$\langle\alpha|\,\hat{X}_1\,|\alpha\rangle = 2Re[\alpha] \tag{3.5}$$
$$\langle\alpha|\,\hat{X}_2\,|\alpha\rangle = 2Im[\alpha] \tag{3.6}$$
$$\langle\alpha|\,\hat{n}\,|\alpha\rangle = |\alpha|^2 \tag{3.7}$$
$$|\,\langle n|\alpha\rangle\,|^2 = \frac{|\alpha|^{2n}}{n!}e^{-|\alpha|^2} \tag{3.8}$$
$$\langle\beta|\alpha\rangle = \exp[-(|\alpha|^2+|\beta|^2)/2+\alpha\beta^*] \tag{3.9}$$
$$|\,\langle\beta|\alpha\rangle\,|^2 = \exp[-|\alpha-\beta|^2] \tag{3.10}$$
$$\frac{1}{\pi}\int d^2\alpha\,|\alpha\rangle\,\langle\alpha| = 1 \tag{3.11}$$
$$\tag{3.12}$$

Like the vacuum state, coherent states are minimum uncertainty states,

$$(\Delta\hat{X}_{1,2})^2 = \langle\alpha|\,\hat{X}_{1,2}^2\,|\alpha\rangle - \langle\alpha|\,\hat{X}_{1,2}\,|\alpha\rangle^2 = 1. \tag{3.13}$$

In fact, some authors prefer to define the coherent states as the ground state displaced to finite $< \hat{X}_{1,2} >$, as $|\alpha\rangle \equiv D(\alpha)\,|0\rangle$ where D is the displacement operator $D(\alpha) \equiv \exp[\alpha\hat{a}^\dagger - \alpha^*\hat{a}]$.

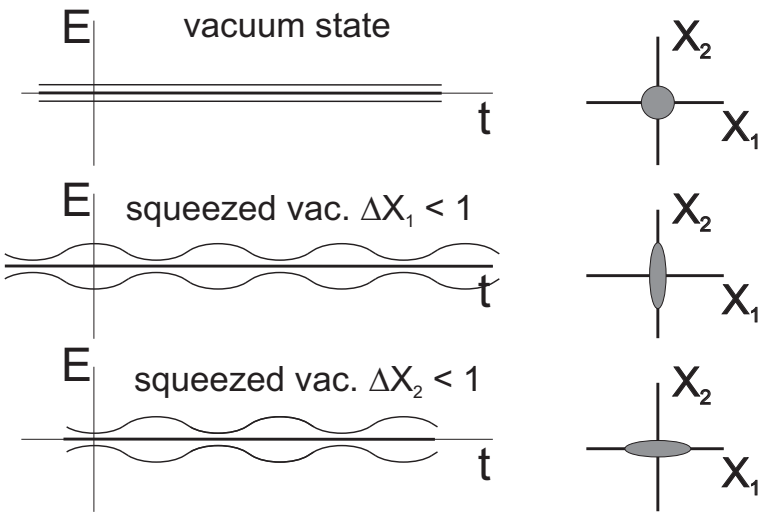

Figure 3.2: Left plots: Average value and uncertainty for a vacuum state (top) and two "squeezed vacuum" states with the same average values for the field. In each plot, the heavy line indicates the average field value $< E(t) >= 0$ while the light lines indicate the average plus/minus $\Delta E(t)$. Right plots: uncertainty ellipse representations of these states.

Squeezed states

To introduce squeezed states, we look a bit at what exactly the uncertainty relation between X_1 and X_2 implies. The average field at one point in space oscillates as

$$\langle E(t) \rangle \propto \left\langle \hat{X}_1 \right\rangle \sin(\omega t) - \left\langle \hat{X}_2 \right\rangle \cos(\omega t). \tag{3.14}$$

At the same time, the variance in the field oscillates as

$$(\Delta E(t))^2 \propto (\Delta \hat{X}_1)^2 \sin^2(\omega t) + (\Delta \hat{X}_2)^2 \cos^2(\omega t) - 2 \sin(\omega t) \cos(\omega t) \mathrm{cov}(\hat{X}_1, \hat{X}_2) \tag{3.15}$$

where the covariance $\mathrm{cov}(\hat{X}_1, \hat{X}_2) \equiv (< \hat{X}_1 \hat{X}_2 > + < \hat{X}_2 \hat{X}_1 >)/2 - < \hat{X}_2 >< \hat{X}_1 >$ reflects the degree of correlation of X_1 and X_2. For the states we consider, the variation of the fields is uncorrelated, $\mathrm{cov}(\hat{X}_1, \hat{X}_2) = 0$. It is clear that the uncertainty relation between X_1 and X_2 implies an uncertainty between $E(t = 0)$ and $E(t = \pi/2\omega)$. Also, we note that the variance $(\Delta E(t))^2$ oscillates at 2ω, while the field itself oscillates at ω. To give a concrete example, we consider possible minimum-uncertainty states with $< \hat{X}_1 >= 6$ and $< \hat{X}_2 >= 0$. If $\Delta X_1 = \Delta X_2 = 1$ we have a coherent state ($\alpha = 3 + 0\,i$). If $\Delta X_1 < 1$ or $\Delta X_2 < 1$ we have a quadrature-squeezed state. The field as a function of time for these states is represented Figure 3.1. Note that for $\Delta X_1 < 1$ the amplitude of oscillation is better defined than for the coherent state, and for $\Delta X_2 < 1$ the zero-crossing is better defined. This may be the origin of the terms "amplitude quadrature" for $\hat{X}_1 \equiv \hat{a} + \hat{a}^\dagger$ and "phase quadrature" for $\hat{X}_2 \equiv i(\hat{a}^\dagger - \hat{a})$.

It is also possible to "squeeze" the fluctuations associated with the vacuum state. A state with zero average $< \hat{X}_1 >=< \hat{X}_2 >= 0$ and reduced fluctuations on one quadrature is called "squeezed vacuum." This is illustrated in Figure 3.2.

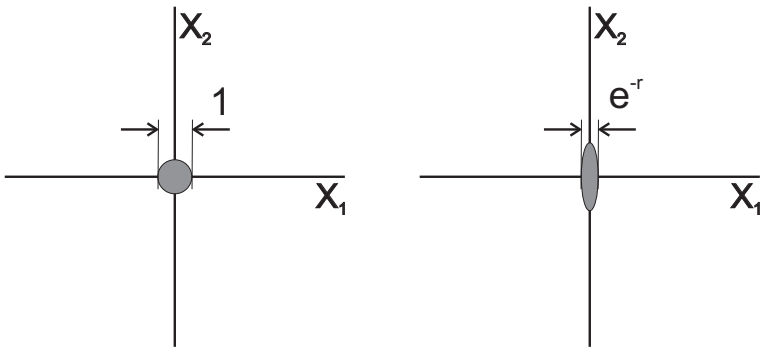

Figure 3.3: Uncertainty ellipse for a vacuum state (left) and squeezed vacuum with $\Delta X_1 < 1$ (right).

Squeezed states are closely related to coherent states. For one thing, we say a state is "squeezed" if it is lower noise than a coherent state. Specifically, one calculates the noise level using a coherent state and defines this as the "standard quantum limit[2]." If the noise is lower than the standard quantum limit, we say the state is squeezed. This can be applied to any measurable quantity. What we just described are quadrature-squeezed states, because one quadrature is better defined, i.e., has lower variance, than the standard quantum limit $\Delta X_{1,2} = 1$. There are also "number-squeezed" states, with $\Delta n < \sqrt{< n >}$, "phase-squeezed" states with $\Delta\phi < 1/\sqrt{< n >}$, and others.

Squeezed states can be generated from the vacuum state by applying the squeeze operator

$$S(\varepsilon) \equiv \exp[\frac{1}{2}\varepsilon^* \hat{a}^2 - \frac{1}{2}\varepsilon(\hat{a}^\dagger)^2]. \tag{3.16}$$

In general, the parameter $\varepsilon = r\exp[2i\phi]$ is complex, and the following useful relations hold

$$\begin{aligned}
S^\dagger(\varepsilon) &= S^{-1}(\varepsilon) = S(-\varepsilon) && (3.17)\\
S^\dagger(\varepsilon)\hat{a}S(\varepsilon) &= \hat{a}\cosh r - \hat{a}^\dagger e^{2i\phi}\sinh r && (3.18)\\
S^\dagger(\varepsilon)\hat{a}^\dagger S(\varepsilon) &= \hat{a}^\dagger\cosh r - \hat{a}e^{-2i\phi}\sinh r && (3.19)\\
S^\dagger(\varepsilon)Y_1 S(\varepsilon) &= Y_1 e^{-r} && (3.20)\\
S^\dagger(\varepsilon)Y_2 S(\varepsilon) &= Y_2 e^{r} && (3.21)
\end{aligned}$$

where $Y_1 \equiv ae^{-i\phi} + a^\dagger e^{i\phi}$, $Y_2 \equiv i(a^\dagger e^{i\phi} - ae^{-i\phi})$ are rotated quadrature operators. When $\phi = 0$, we have

$$\begin{aligned}
S^\dagger(r)X_1 S(r) &= X_1 e^{-r} && (3.22)\\
S^\dagger(r)X_2 S(r) &= X_2 e^{r}. && (3.23)
\end{aligned}$$

[2]The name "standard quantum limit" may appear strange. From the perspective of quantum theory is not a limit at all. It is the value one gets when a particular state (a coherent state) is used. The name comes from experiment, in which there are always other noise sources which the experimenter must try to eliminate. With a coherent state, these efforts can only reduce the noise to the coherent state noise level. If the noise level drops below this limit, it is experimental proof that the state was squeezed.

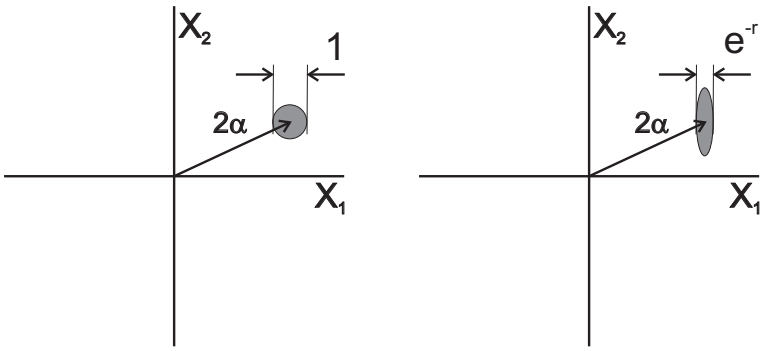

Figure 3.4: Uncertainty areas of a coherent state (left) and a bright squeezed state with $\Delta X_1 < 1$ (right).

Evidently squeezing a state reduces its amplitude quadrature by a factor of $\exp[r]$ while increasing its phase quadrature by the same amount. The state $S(\varepsilon)\,|0\rangle$ is called "squeezed vacuum." A convenient way to represent such states pictorially is in terms of their "uncertainty ellipses" or "error ellipses" in the X_1, X_2 plane. Two such diagrams are shown in Figure 3.3.

Squeezed states which have non-zero average fields can be produced by applying the squeeze operator and then the displacement operator to the vacuum state, as $|\alpha, \varepsilon\rangle \equiv D(\alpha)S(\varepsilon)\,|0\rangle$. These states are sometimes called "bright squeezed states" or in the laboratory "bright squeezed beams." These are shown in Figure 3.4. They have the following properties

$$\langle X_1 + iX_2\rangle = 2\alpha \tag{3.24}$$

$$\langle N\rangle = |\alpha|^2 + \sinh^2 r \tag{3.25}$$

$$(\Delta N)^2 = |\alpha\cosh r - \alpha^* e^{2i\phi}\sinh r|^2 + 2\cosh^2 r\sinh^2 r. \tag{3.26}$$

Note that squeezing the vacuum adds some photons to the field, as shown by equation 3.25. This means that "squeezed vacuum" contains a small but nonzero flux of photons.

3.5 Entangled states

Entanglement is fairly easy to generate in quantum optics. How this is done will be described later, here we just note that this is one of the main reasons for the current interest in quantum optics for quantum information. We first consider the case for photon-counting, using number states, then with quadrature states.

Entanglement necessarily involves multiple quantum systems. They could be multiple photons or multiple modes.

Consider the state

$$|DA\rangle = \frac{1}{2}(\hat{a}^\dagger_{H1} + \hat{a}^\dagger_{V1})(\hat{a}^\dagger_{H2} - \hat{a}^\dagger_{V2})\,|0\rangle \tag{3.27}$$

where $H1, V1, H2, V2$ are four distinct modes describing horizontal (H) and vertical (V) polarization for two distinct modes 1,2. Because the combination $(\hat{a}^\dagger_{H2} - \hat{a}^\dagger_{V2})/\sqrt{2}$ creates a single photon, we can interpret this as a the creation operator $\hat{a}^\dagger_{A2}$ for a photon with polarization $A \equiv (H - V)/\sqrt{2}$. Similarly the first photon is created by the creation operator $(\hat{a}^\dagger_{H1} + \hat{a}^\dagger_{V1})/\sqrt{2} = \hat{a}^\dagger_{D1}$ where $D \equiv (H + V)/\sqrt{2}$. Thus the state can be re-written as

$$|DA\rangle = \hat{a}^\dagger_{D1}\hat{a}^\dagger_{A2}\,|0\rangle\,. \tag{3.28}$$

This state simply describes two photons in two different modes, each with a different polarization. If we write this the way it would be written in ordinary quantum mechanics, it would be

$$|DA\rangle = |D\rangle_1\,|A\rangle_2 \tag{3.29}$$

where $|\phi_{1,2}\rangle_{1,2}$ is the state of the 1st or 2nd photon. This is a "product state." In contrast, the state

$$\left|\Psi^-\right\rangle = \frac{1}{\sqrt{2}}(\hat{a}^\dagger_{H1}\hat{a}^\dagger_{V2} - \hat{a}^\dagger_{V1}\hat{a}^\dagger_{H2})\,|0\rangle = \frac{1}{\sqrt{2}}(|H\rangle_1\,|V\rangle_2 - |V\rangle_1\,|H\rangle_2) \tag{3.30}$$

cannot be factorized (written as a product) and is thus entangled. In fact, the state Ψ^- is called a "Bell state" and it is often discussed in connection with quantum nonlocality and the violation of Bell inequalities.

The above example shows entanglement in polarization, a discrete variable described in terms of just two states H, V. Entanglement in continuous variables such as quadratures is also possible. The best known example of this is the Einstein-Podolsky-Rosen (EPR) paradox, in which two particles have correlated positions $x_1 - x_2 =$ const. and anti-correlated momenta $p_1 + p_2 = 0$. The individual particles' position and momentum are completely uncertain, it is only the relative coordinate and combined momentum that are sharp. The EPR situation can not be described by a product state of a wave-function for particle 1 times a wave function for particle 2. More generally, it was shown by Duan, Giedke, Cirac and Zoller in 1999 that when the correlated variances are sufficiently small $(\Delta(X_A - X_B))^2 + (\Delta(P_A + P_B))^2 < 2$, the state must be entangled, i.e., not factorizable. Here $\hat{X}, \hat{P}$ are scaled variables with the commutation relation $[\hat{X}, \hat{P}] = i$.

It turns out that a state with EPR correlations in the quadratures of two different modes is also fairly easy to make in quantum optics. Again, we will show how to do this later, and for the moment we just show what such a state would look like. Consider the vacuum state $|0\rangle$ of two different modes at frequencies ω_+, ω_-. Now squeeze this state using the unitary two-mode squeeze operator $S_2(G) = \exp[G^*\hat{a}_+\hat{a}_- - G\hat{a}^\dagger_+\hat{a}^\dagger_-]$. The squeeze operator transforms the annihilation operators as

$$S_2^\dagger(G)\hat{a}_\pm S_2(G) = \hat{a}_\pm \cosh r - \hat{a}^\dagger_\mp e^{i\theta}\sinh r. \tag{3.31}$$

To keep things simple, we take $G = r\exp[i\theta]$ to be real, i.e. $\theta = 0$. We define the sum and difference quadratures

$$\hat{X}_{1s} \equiv (\hat{X}_{1+} + \hat{X}_{1-})/\sqrt{2} \tag{3.32}$$

$$\hat{X}_{2d} \equiv (\hat{X}_{2+} - \hat{X}_{2-})/\sqrt{2} \tag{3.33}$$

and with a bit of algebra it can be shown that for the state $S_2(r)\,|0\rangle$,

$$(\Delta\hat{X}_{1s})^2 = e^{-2r} \tag{3.34}$$
$$(\Delta\hat{X}_{2d})^2 = e^{-2r}. \tag{3.35}$$

This shows that it is possible to have a state of two modes which is squeezed in the sum of the amplitude quadratures and also in the difference of the phase quadratures. This same state is anti-squeezed (variance larger than the coherent state value) for the difference of the amplitude quadratures and the sum of phase quadratures. A state like this can be used to demonstrate continuous-variable entanglement by violating the inequality given by Duan et al. above.

This ends our sampling of the possible quantum states, but we have not exhausted the possibilities. In fact the number of possible states grows exponentially with the number of photons (or the number of modes) available. Thus there are an infinitude of different states, and most of them have large numbers of photons and are not close to classical states. In a sense, quantum optics is still just scratching the surface of the available quantum states. As experiments in optical quantum information advance, the states we use will become more and more entangled, and less and less classical. Maybe some day the term "classical optics" will describe the unusual situation, rarely encountered, of an experiment that uses only coherent states.

Chapter 4

Detection of light

There are two principal ways of detecting light that are used in quantum optics. One, "direct detection," detects the energy falling on a detector, and is closely related to the number-state basis, because this is the energy basis. The other method is to mix the signal beam with a strong reference of the same wavelength and definite phase. The interference is detected as a power difference at the outputs of the beam-splitter. This depends on the phase of the measured beam, and the result is detection of a single quadrature. Naturally, such experiments are best explained using quadratures.

4.1 Direct detection and photon counting

The simplest method of detecting light, called "direct detection," is to absorb the light on the surface of a detector of some sort (a photodiode, a photomultiplier tube, a thermal detector, etc.). The detector produces an electrical signal proportional to the power of the incident light. Classically, such a detector is called a "square-law" detector because the electrical signal (voltage or current) is proportional to the square of the incident electric field. Quantum mechanically, the signal indicates the number of photons that have been absorbed by the detector. If the detector is sensitive enough, individual photon arrivals can be observed, and we speak of detection by "photon counting."

The theory of photon counting was first presented by Roy Glauber in 1964 [1]. He noted that while a classical photo-detection signal is proportional to the square of the electric field averaged over a few cycles $P^{(\text{Class.})}(t) \propto \left\langle E^2(t) \right\rangle$, the same can not be true for quantum fields. In particular, because of vacuum fluctuations, $\left\langle \hat{\mathbf{E}}^2(t) \right\rangle > 0$ even for the vacuum state. If we naïvely applied the classical detection formula, it would imply detections even when there are no photons present.

[1] R. J. Glauber, Quantum Optics and Electronics, Les Houches Summer Lectures 1964, edited by C. DeWitt, A. Blandin, and C. Cohen-Tannoudji (Gordon and Breach, New York, 1965)

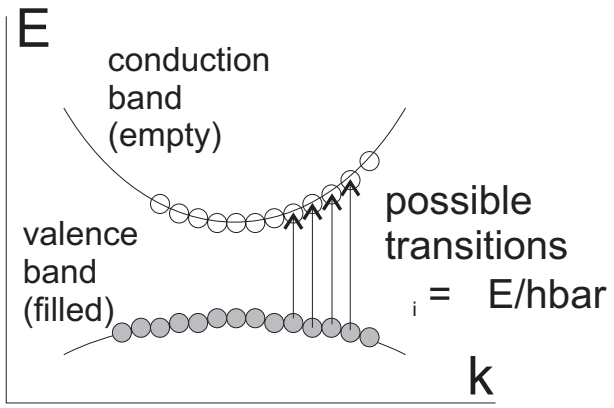

Figure 4.1: Transitions in an idealized semiconductor.

In fact, we will see that the detection rate is given by

$$P(\mathbf{r}, t) \propto \left\langle \hat{\mathbf{E}}^{(-)}(\mathbf{r}, t) \cdot \hat{\mathbf{E}}^{(+)}(\mathbf{r}, t) \right\rangle \tag{4.1}$$

where

$$\hat{\mathbf{E}} = \hat{\mathbf{E}}^{(+)} + \hat{\mathbf{E}}^{(-)} \tag{4.2}$$

and

$$\hat{\mathbf{E}}^{(+)}(\mathbf{r}, t) = i \sum_{k,\alpha} \sqrt{\frac{\hbar\omega_k}{2\varepsilon_0}} \hat{a}_{k,\alpha} \mathbf{u}_{k,\alpha}(\mathbf{r}) e^{-i\omega_k t} \tag{4.3}$$

is called the positive-frequency part of the field and $\hat{\mathbf{E}}^{(-)}(\mathbf{r}, t) = [\hat{\mathbf{E}}^{(+)}(\mathbf{r}, t)]^\dagger$ is called the negative-frequency part of the field. Note that $\hat{\mathbf{E}}^{(+)}$ contains only annihilation operators, so that it acts on the vacuum state to produce zero. Thus Glauber's theory does not predict detections in the absence of photons. We now describe Glauber's argument.

Glauber considered the interaction of the quantized field with a detector consisting of many atoms with different transition frequencies. Here we use the same argument, but apply it to a semiconductor detector such as an avalanche photodiode. As shown in Figure 4.1, we assume a filled valence band containing a very large number of electrons and an empty conduction band. We assume that an electron promoted into the conduction band can be detected efficiently. In fact, for modern avalanche photodiodes this is the case: any free electron is swept into a high-field amplification region, where it is accelerated and creates many electron-hole pairs. Detection of these secondary electrons can then be done by ordinary electronic amplifiers. We thus concern ourselves with just the first step, the promotion of a single electron into the conduction band. We assume that the detector starts in its ground state $|0\rangle_{det} = |v\rangle_1 |v\rangle_2 ...$ with all electrons in the valence band.

The ith electron can be promoted to the valence band by absorbing an energy $\hbar\omega_i$. The dipole matrix element (an operator) for this transition is $\hat{d}_i \equiv d_0(|v\rangle_i \langle c|_i + |c\rangle_i \langle v|_i) = d_0(\hat{b}_i + \hat{b}_i^\dagger)$ where for convenience we have defined $\hat{b} \equiv |v\rangle \langle c|$. We assume the Hamiltonian

$$\hat{H} = \hat{H}_0 + \hat{H}_I \tag{4.4}$$

where

$$\hat{H}_0 = \sum_k \hbar\omega_k(\hat{a}_k^\dagger \hat{a}_k + \frac{1}{2}) + \sum_i \hbar\omega_i \hat{b}_i^\dagger \hat{b}_i \tag{4.5}$$

and

$$\begin{aligned} \hat{H}_I &= -\hat{E}\hat{d} \\ &= -\sum_i \left(d_0 \hat{E}^{(+)} b_i^\dagger + d_0^* \hat{E}^{(-)} b_i \right) - \sum_i \left(d_0 \hat{E}^{(-)} b_i^\dagger + d_0^* \hat{E}^{(+)} b_i \right) \\ &\approx -\sum_i \left(d_0 \hat{E}^{(+)} b_i^\dagger + d_0^* \hat{E}^{(-)} b_i \right). \end{aligned} \tag{4.6}$$

For clarity of presentation we assume just one polarization, and we drop the second sum because it greatly fails to conserve energy. Dropping this kind of term is known as the "rotating-wave approximation" and is discussed in greater xin Chapter 10.

The detection rate for this model is calculated in detail in Appendix B, here we just give an outline. Treating H_I as a perturbation, we first observe that to zero'th order the field $E_0(t)$ evolves under Maxwell's equations from whatever is the initial conditions. Also under zero'th order, the probability of a given electron being excited $\left\langle b_i^\dagger b_i \right\rangle$ and the coherence between valence and conduction states $\langle b_i \rangle$ are zero. In second order time-dependent perturbation theory, however, the probability of excitation grows as (Equation B.20)

$$\frac{d}{dt}\langle n_i(t)\rangle = \frac{|g|^2}{\hbar^2}\int_0^t dt' \left\langle E_0^{(-)}(t) E_0^{(+)}(t') e^{-i\omega_i(t-t')} + E_0^{(-)}(t') E_0^{(+)}(t) e^{i\omega_i(t-t')} \right\rangle \tag{4.7}$$

with $|g|^2 = |d_0|^2$. To get the total number of electrons excited, we must sum n_i over all the electrons, keeping in mind that different electrons have different ω_i. Converting this sum to an integral $\sum_i \to \int d\omega\rho$, where ρ is the density of states[2], we obtain a delta-function $2\pi\rho\delta(t-t')$, so that the rate of excitation becomes

$$\frac{d}{dt}\langle N(t)\rangle \equiv \frac{d}{dt}\sum_i \langle n_i\rangle = 2\pi\rho\frac{|g|^2}{\hbar^2}\left\langle E_0^{(-)}(t) E_0^{(+)}(t) \right\rangle. \tag{4.8}$$

This is Glauber's result. To be clear, E_0 is simply the field that enters the detector, as it would evolve if the detector were not present. In this sense, it is an ideal measurement. On the other hand, the input field is consumed in the measurement, so it is very destructive!

A note of caution: the result above, $P(t) \propto \left\langle \hat{E}^{(-)}(t)\hat{E}^{(+)}(t) \right\rangle$ or more generally $P(x,t) \propto \left\langle \hat{E}^{(-)}(x,t)\hat{E}^{(+)}(x,t) \right\rangle$ is used, implicitly or explicitly, to explain almost all photon-counting experiments. In contrast, the proportionality constant $2\pi\rho\left|\frac{d_0}{\hbar}\right|^2$ is almost never used. In fact, it is not correct for most photon detectors. The expression was derived by perturbation theory,

[2]The density of states in fact depends on frequency ω, but the dependence is not very strong: $d\rho/d\omega \sim \rho/\omega$, whereas $\Delta\omega$, the range of frequencies of interest to us, is only the spectral width of the light we are detecting. This is typically very narrow, $\Delta\omega \ll \omega$, so that $\Delta\rho \ll \rho$. For this reason, we can take the density of states as approximately constant.

assuming that the probability of absorbing a photon was small. But most optical detectors are highly opaque to incident light. Roughly speaking, the expression with $2\pi\rho\left|\frac{d_0}{\hbar}\right|^2$ is the absorption probability in the first thin slice of the detector, and the absorption probability decays exponentially with depth beneath the surface. For an opaque, efficient detector, the sum of all the layers is one detection per incident photon. Typically we keep the result $P(x,t) \propto \left\langle \hat{E}^{(-)}(x,t)\hat{E}^{(+)}(x,t)\right\rangle$, and find some other way to determine the absolute rate of detections. For example, if somehow we know that the average power falling on the detector is P_{opt}, then the average rate of detection is $P_{opt}/\hbar\omega$.

Coincidence counting

The expression above describes the detection probability for a single detector. What if there are multiple detectors (almost always the case in photon counting experiments)? Then we may be interested in correlations among the detections, for example, "if detector A fires, so does detector B" or "detector A *never* fires exactly one nanosecond after detector B."

Glauber also considered this situation. To see if *two* electrons have been excited, one in each of detectors A, B, we calculate the evolution of the operator

$$\hat{N}_2(t_A, t_B) \equiv \sum_{i,j} \hat{b}_i^\dagger(t_A)\hat{b}_j^\dagger(t_B)\hat{b}_j(t_B)\hat{b}_i(t_A) \tag{4.9}$$

where the $\hat{b}_i$, $\hat{b}_j$ act on electrons in detectors A,B. The probability density of seeing two detections at times t_A, t_B is

$$P(t_A, t_B) = \frac{\partial^2}{\partial t_A \partial t_B}\left\langle \hat{N}_2(t_A, t_B)\right\rangle \tag{4.10}$$

and the perturbation calculation finds

$$P(t_A, t_B) \propto \left\langle \hat{E}^{(-)}(\mathbf{r}_A, t_A)\hat{E}^{(-)}(\mathbf{r}_B, t_B)\hat{E}^{(+)}(\mathbf{r}_B, t_B)\hat{E}^{(+)}(\mathbf{r}_A, t_A)\right\rangle. \tag{4.11}$$

Note that the order of the operators is important. All of the annihilation operators are on the right, so a state with insufficient photons (fewer than two in this case) is annihilated. The extension to N-photon detection is obvious.

4.2 Homodyne detection

Photon counting necessarily detects intensities or powers. Is it possible to detect the field amplitude somehow? In principle, an electromagnetic field is observable with an antenna and a sufficiently fast oscilloscope. But for optical fields the oscillations are too rapid and the wavelength is too short, so we have to find other ways. A very elegant technique is *balanced homodyne detection*, as illustrated in Figure 4.2. A beamsplitter is used to mix the field we want to measure, the "signal" with a strong reference beam, the "local oscillator" (LO) whose phase ϕ

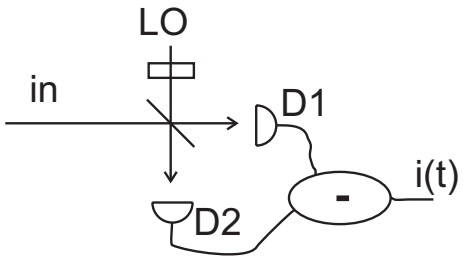

Figure 4.2: Homodyne detection.

we can control. We assume that the beamsplitter is *balanced* meaning that the transmission and reflection coefficients are equal magnitude. We also assume that the conditions for interference are ideal: the LO is in a single spatial mode, is monochromatic, and has a constant phase. Naturally, the input field must be matched to this mode. At each output port of the beamsplitter, a detector $D1$ or $D2$ detects all the light that leaves by that port. The photocurrents from these two detectors are immediately subtracted, so that the output signal is $\Delta i(t) \propto P_1(t) - P_2(t)$ where $P_{1,2}(t)$ are the powers arriving at detectors D1 and D2.

We analyze the situation classically first. The LO field is E_{LO}, the input field is E_{in}. They are assumed to have the same optical frequency ω. The fields leaving the beamsplitter are

$$E_1 = \frac{1}{\sqrt{2}}(E_{LO} + E_{in}) \tag{4.12}$$

$$E_2 = \frac{1}{\sqrt{2}}(E_{LO} - E_{in}). \tag{4.13}$$

The detected powers are

$$P_1 \propto \left\langle E_1^2 \right\rangle = \frac{1}{2}(\left\langle E_{LO}^2 \right\rangle + \left\langle E_{in}^2 \right\rangle + 2\left\langle E_{LO}E_{in} \right\rangle) \tag{4.14}$$

$$P_2 \propto \left\langle E_2^2 \right\rangle = \frac{1}{2}(\left\langle E_{LO}^2 \right\rangle + \left\langle E_{in}^2 \right\rangle - 2\left\langle E_{LO}E_{in} \right\rangle) \tag{4.15}$$

where the brackets indicate time-averaging over several optical cycles. The subtraction of the signals gives

$$\Delta i(t) \propto \left\langle E_{LO}(t)E_{in}(t) \right\rangle. \tag{4.16}$$

It is clear already that this technique should be useful for the detection of weak fields: the signal strength is proportional to $\langle E_{LO}(t)E_{in}(t)\rangle$, much larger than than the signal strength with direct detection, proportional to $\langle E_{in}(t)E_{in}(t)\rangle$. Furthermore, in terms of quadratures, $E_{LO}(t) = X_1^{(LO)} \sin\omega t - X_2^{(LO)} \cos\omega t$ and $E_{in}(t) = X_1^{(in)} \sin\omega t - X_2^{(in)} \cos\omega t$, we have

$$\Delta i(t) \propto X_1^{(LO)} X_1^{(in)} + X_2^{(LO)} X_2^{(in)}. \tag{4.17}$$

Represented in terms of phasors $2\alpha = X_1 + iX_2$ (these will later become coherent state amplitudes), we find that

$$\Delta i(t) \propto Re[\alpha_{LO}^* \alpha_{in}]. \tag{4.18}$$

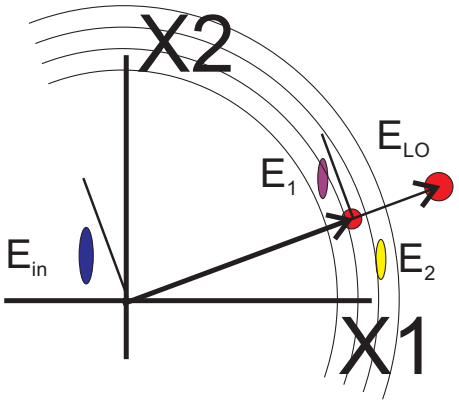

Figure 4.3: Phase-space representation of quadrature detection. A strong local oscillator field is mixed with the input signal, giving $E_1 = (E_{LO} + E_{in})/\sqrt{2}$ and $E_2 = (E_{LO} - E_{in})/\sqrt{2}$. Curved lines show contours of constant power. The detected signal $\Delta i \propto < E_1^2 > - < E_2^2 >$ is a measure of one generalized quadrature of E_{in}, the one in-phase with E_{LO}.

We note a few very attractive features of this measurement technique. As mentioned already, this offers a way to boost weak signals, by mixing them with a strong reference. This is the basis of most techniques in radio transmission, for example. It also allows us to make quadrature measurements. For example, if we choose α_{LO} to be real, so that $X_2^{(LO)} = 0$, then the signal indicates only the real part of α_{in}, or equivalently only the quadrature $X_1^{(in)}$. By changing the phase ϕ of the LO, we can measure $X_1^{(in)}$, $X_2^{(in)}$, or any linear combination (a generalized quadrature) $X_1^{(in)} \sin\phi + X_2^{(in)} \cos\phi$. Finally, we note that the technique is very favorable for low-noise measurements. Noise in the LO, for example if $\alpha_{LO} = \alpha_0 + \delta\alpha$, then $\delta\alpha$ contributes to the noise in the signal as $\delta\Delta i(t) \propto Re[\delta\alpha^* \alpha_{in}]$. Because α_{in} is small, noise in the LO has a small effect on the measurement noise. In the words of Hans Bachor, "This is an extremely useful and somewhat magical device."

A pictorial representation of the homodyne measurement process is shown in Figure 4.3.

The quantum mechanical description of homodyne measurement is very simple, but assumes a quantum-mechanical understanding of beamsplitters that we will develop later. The result of that understanding is that the beamsplitter transforms the quantum fields as

$$\hat{E}_1 = \frac{1}{\sqrt{2}}(\hat{E}_{LO} + \hat{E}_{in}) \tag{4.19}$$

$$\hat{E}_2 = \frac{1}{\sqrt{2}}(\hat{E}_{LO} - \hat{E}_{in}). \tag{4.20}$$

In other words, the quantum beamsplitter acts just like the classical one. The detection process, treated quantum mechanically, gives the same results as the classical treatment because when detected each beam contains many photons and is nearly classical. Following Glauber, we would

write each photocurrent as

$$i_1 \propto \left\langle \hat{E}_1^{(-)} \hat{E}_1^{(+)} \right\rangle = \left\langle \hat{E}_{LO}^{(-)} \hat{E}_{LO}^{(+)} + \hat{E}_{in}^{(-)} \hat{E}_{in}^{(+)} + \hat{E}_{LO}^{(-)} \hat{E}_{in}^{(+)} + \hat{E}_{in}^{(-)} \hat{E}_{LO}^{(+)} \right\rangle \tag{4.21}$$

$$i_2 \propto \left\langle \hat{E}_2^{(-)} \hat{E}_2^{(+)} \right\rangle = \left\langle \hat{E}_{LO}^{(-)} \hat{E}_{LO}^{(+)} + \hat{E}_{in}^{(-)} \hat{E}_{in}^{(+)} - \hat{E}_{LO}^{(-)} \hat{E}_{in}^{(+)} - \hat{E}_{in}^{(-)} \hat{E}_{LO}^{(+)} \right\rangle \tag{4.22}$$

so that

$$\Delta i = i_1 - i_2 \propto \left\langle \hat{E}_{LO}^{(-)} \hat{E}_{in}^{(+)} + \hat{E}_{in}^{(-)} \hat{E}_{LO}^{(+)} \right\rangle \propto \hat{X}_1^{(LO)} \hat{X}_1^{(im)} + \hat{X}_2^{(LO)} \hat{X}_2^{(im)}. \tag{4.23}$$

Note that to get this last expression we have used the fact that the LO field is single mode, so that $\hat{E}^{(+)} \propto \hat{a}_k = (\hat{X}_1 + i\hat{X}_2)/2$ and that quadrature operators for the LO and input fields commute $[\hat{X}_1^{(LO)}, \hat{X}_2^{(im)}] = 0$. This gives us the same result as the classical case, but now with the quantized quadrature operators. The rest of the discussion, about noise contributions, signal strengths, etc. is the same.

Chapter 5

Correlation functions

Because many things that we measure in quantum optics are random (quantum noise, photon arrival times from stochastic sources, as well as ordinary noise from imperfect instruments or environmental conditions), we often rely upon correlation functions to describe our results.

Classically, a correlation function is simply the average of a product of two or more quantities, for example the amplitude autocorrelation function is

$$G^{(1)}(\tau) \equiv \langle E(t)E(t+\tau)\rangle = \lim_{T\to\infty} \frac{1}{T}\int_0^T dt\, E(t)E(t+\tau) \tag{5.1}$$

and the amplitude cross-correlation function between fields E_A and E_B is

$$G^{(1)}_{A,B}(\tau) \equiv \langle E_A(t)E_B(t+\tau)\rangle = \lim_{T\to\infty} \frac{1}{T}\int_0^T dt\, E_A(t)E_B(t+\tau). \tag{5.2}$$

Correlation functions are, in general, expressions of the degree of coherence within a single source or between different sources. We illustrate by considering interference between two sources $E_A(t), E_B(t)$ which we combine on a beamsplitter to produce the fields $E_{1,2}(t) \equiv [E_A(t) \pm E_B(t+\tau)]/\sqrt{2}$. Here τ is a small variable delay that we can use to change the relative phase of the fields. After the beamsplitter the fields are detected, giving currents

$$i_{1,2}(\tau) \propto \left\langle [E_A(t) \pm E_B(t+\tau)]^2 \right\rangle /2 = \left\langle E_A^2 \right\rangle /2 + \left\langle E_B^2 \right\rangle /2 \pm \langle E_A(t)E_B(t+\tau)\rangle \tag{5.3}$$

or

$$i_{1,2}(\tau) \propto \left\langle E_A^2 \right\rangle + \left\langle E_B^2 \right\rangle \pm 2G^{(1)}_{A,B}(\tau). \tag{5.4}$$

Note that the interference signal comes entirely from the correlation function $G^{(1)}_{A,B}(\tau)$.

The autocorrelation function $G^{(1)}(\tau)$ above is closely related to spectroscopy. We illustrate with an unbalanced Mach-Zehnder interferometer. The input field $E(t)$ is split into two beams which travel paths which differ in length by $c\tau$. The beams are then combined on a beamsplitter to produce the fields $E_{1,2}(t) \equiv [E(t) \pm E(t+\tau)]/2$. These are detected, giving currents

$$i_{1,2} \propto \left\langle E_{1,2}^2(t) \right\rangle = \left\langle [E(t) \pm E(t+\tau)]^2 \right\rangle /4 = \left\langle E^2 \right\rangle /2 \pm \langle E(t)E(t+\tau)\rangle /2. \tag{5.5}$$

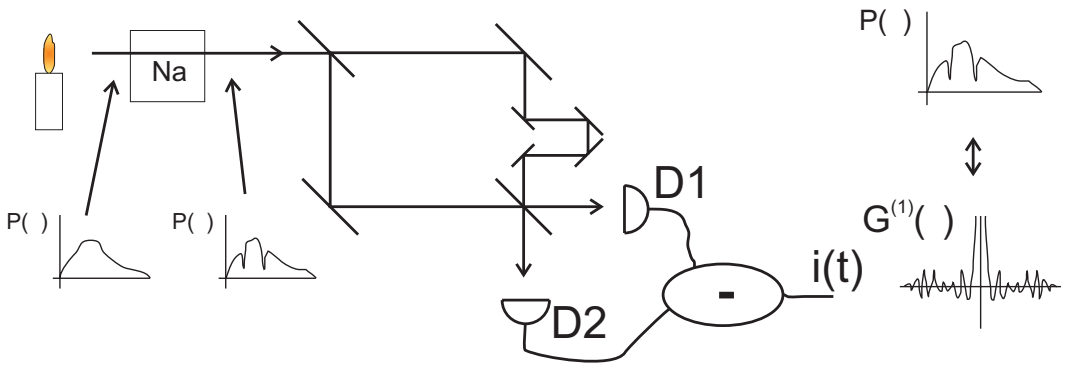

Figure 5.1: Mach-Zehnder interferometer as a spectrometer.

Subtraction of the currents gives $i_- \equiv i_1 - i_2 \propto \langle E(t)E(t+\tau)\rangle = G^{(1)}(\tau)$. Thus the amplitude correlation function is simply the interference signal that we get from the interferometer. At the same time, it contains the spectrum of the input light. To see this, we note the correlation theorem from Fourier theory

$$G^*(\nu)H(\nu) \leftrightarrow \int_{-\infty}^{\infty} dt\, g(t)h(t+\tau). \tag{5.6}$$

Here the symbol $\leftrightarrow$ indicates Fourier transform and $G(\nu) \leftrightarrow g(\tau), H(\nu) \leftrightarrow h(\tau)$. When applied with $g(t) = h(t) = E(t)$, this immediately yields

$$|E(\nu)|^2 \leftrightarrow G^{(1)}(\tau). \tag{5.7}$$

In words, the spectrum is the Fourier transform of the amplitude auto-correlation function.

5.1 Quantum correlation functions

Quantum mechanical correlation functions are analogous to the classical versions, with two important differences. First, we replace the classical fields E with quantum field operators, which could be $\hat{X}_{1,2}$, $\hat{E}^{(+)}$ or $\hat{E}^{(-)}$. For example, in the spectroscopy example above, the quantum version of the amplitude autocorrelation function is

$$G^{(1)}(\tau) \equiv \left\langle \hat{E}^{(-)}(t)\hat{E}^{(+)}(t+\tau)\right\rangle. \tag{5.8}$$

Second, we interpret the averaging brackets $\langle\rangle$ as an expectation value, with the state of the field given either by a pure state $|\phi\rangle$,

$$G^{(1)}(\tau) = \langle\phi|\, \hat{E}^{(-)}(t)\hat{E}^{(+)}(t+\tau)\,|\phi\rangle \tag{5.9}$$

or a density matrix ρ

$$G^{(1)}(\tau) = \mathrm{Tr}[\rho\hat{E}^{(-)}(t)\hat{E}^{(+)}(t+\tau)]. \tag{5.10}$$

Note that in many cases the averaging brackets imply *both* an expectation value and a time average. This is the case in the above expressions, where the average over t is implied by the fact that $G^{(1)}(\tau)$ does not contain t. As an example of the other sort, recall that in Glauber's photodetection theory the probability density of detecting a photon at time t was

$$P(t) \propto \left\langle \hat{E}^{(-)}(t)\hat{E}^{(+)}(t) \right\rangle . \tag{5.11}$$

5.2 Intensity correlations

As noted already, field correlation functions such as $G^{(1)}(\tau)$ are important in classical optics for describing partial coherence. In contrast, intensity correlation functions appear much less commonly. Nevertheless, they have been important in astronomy, where R. Hanbury-Brown was able to measure the diameters of stars using intensity correlations in radio signals.

Classically, an intensity cross-correlation function between two signals A and B is

$$G^{(2)}_{A,B}(\tau) \equiv \langle I_A(t) I_B(t+\tau) \rangle . \tag{5.12}$$

If the two sources are correlated, then $G^{(2)}_{A,B}(0) > \langle I_A \rangle \langle I_B \rangle$, while if they are uncorrelated, $G^{(2)}_{A,B}(0) = \langle I_A \rangle \langle I_B \rangle$. Hanbury-Brown used two radio-telescopes pointed to the same star to collect the intensities I_A, I_B. When the telescopes were sufficiently close to each other, i.e., within a coherence length, the intensities were strongly correlated. When they were separated by more than a coherence length, the correlations dropped off. This way Hanbury-Brown was able to measure the coherence length and thus the angular size of the stars. Practically, it was much easier to measure $G^{(2)}_{A,B}(0)$ than an amplitude correlation function, because there was no need to preserve the phase of the rapidly-varying radio fields. It was sufficient to detect and multiply intensities, which were relatively slowly varying.

Intensity correlations play a very important role in quantum optics, especially in photon-counting experiments. From Glauber's theory, a product of four operators describes the probability density for coincidence detection of two photons

$$P(t_A, t_B) \propto \left\langle \hat{E}^{(-)}_A(t_A)\hat{E}^{(-)}_B(t_B)\hat{E}^{(+)}_B(t_B)\hat{E}^{(+)}_A(t_A) \right\rangle . \tag{5.13}$$

If we define $t_A \equiv t$ and $t_B \equiv t + \tau$ and average this expression over t we get the probability for seeing a pair of detections separated by a time τ

$$G^{(2)}_{A,B}(\tau) \equiv \left\langle \hat{E}^{(-)}_A(t)\hat{E}^{(-)}_B(t+\tau)\hat{E}^{(+)}_B(t+\tau)\hat{E}^{(+)}_A(t) \right\rangle . \tag{5.14}$$

A special case is when A and B are copies of the same field, for example if a single beam is split to two detectors by a beamsplitter. Then we have

$$G^{(2)}(\tau) \equiv \left\langle \hat{E}^{(-)}(t)\hat{E}^{(-)}(t+\tau)\hat{E}^{(+)}(t+\tau)\hat{E}^{(+)}(t) \right\rangle . \tag{5.15}$$

Finally, we note that the various G functions we have written all have units of some sort. It is often convenient to work with normalized correlation functions, for example

$$g^{(2)}(\tau) \equiv \frac{\left\langle \hat{E}^{(-)}(t)\hat{E}^{(-)}(t+\tau)\hat{E}^{(+)}(t+\tau)\hat{E}^{(+)}(t)\right\rangle}{\left\langle \hat{E}^{(-)}(t)\hat{E}^{(+)}(t)\right\rangle^2} = \frac{G^{(2)}(\tau)}{\langle I\rangle^2}. \tag{5.16}$$

This last function, $g^{(2)}(\tau)$, appears in so many important experiments, it can be called "gee-2" without risk of confusion.

5.3 Measuring correlation functions

Measuring correlation functions in the laboratory is straightforward. We consider as an example the measurement of $g^{(2)}(0)$ and distinguish a couple of measurement scenarios. If the detectors are unable to resolve individual photon arrivals, either because there are too many, or because the detector noise is too large, then we must consider the signals to be continuous. The detectors produce photocurrents $i_{1,2}(t) \propto I_{1,2}(t)$ (plus detector noise). Analog electronic circuits are then used to delay i_1, multiply $i_1 \times i_2$, and average the product to obtain a signal proportional to $\langle I_1(t)I_2(t+\tau)\rangle$. If the noise in the two detectors is uncorrelated, it makes no contribution to this average. The individual intensities $\langle I_1(t)\rangle, \langle I_2(t)\rangle$ in $g^{(2)}$ usually do not need to be measured directly. It is almost always the case that $I_1(t)$ and $I_2(t+\tau)$ are uncorrelated for suitably large τ. In this case, $\langle I_1(t)I_2(t+\tau)\rangle \to \langle I_1\rangle\langle I_2\rangle$. Alternately, each photocurrent can be recorded with a fast oscilloscope and the correlation functions computed afterward.

In the case where single-photon counting is used, we have to make allowance for the fact that the signals are discrete: the photons arrive at times t_1, t_2, etc. In principle we could describe the power $P(t)$ reaching the detector as a series of delta functions $P(t) = AI(t) = \hbar\omega[\delta(t-t_1)+\delta(t-t_2)+...]$, where A is the area of the detector. But delta functions are not what we measure in the laboratory, since we never have infinite time-resolution in our measurements. Instead, we divide the time into intervals, called "bins," of duration δt, i.e., $b_k : k\delta t \le t < (k+1)\delta t$. The experimental signal is the number of detections in each time bin, n_k, proportional to the integrated power $n_k = \int_{t\in b_k} dt\, P(t)/\hbar\omega$. Our best estimate of the intensity is "coarse-grained": $I(t) \propto n_i, i = \lfloor t/\delta t\rfloor$. The integrals in the correlation function now become sums, for example

$$\langle I(t)I(t+j\delta t)\rangle = \frac{1}{T}\int_0^T dt\, I(t)I(t+j\delta t) = \frac{1}{N}\frac{\hbar^2\omega^2}{A^2}\sum_i^N n_i n_{i+j} \propto \langle n_i n_{i+j}\rangle. \tag{5.17}$$

As with continuous signals, one strategy is to simply record the detector output. Each time a photon arrives the time of the detector's firing is recorded, so that $n_i = 1$ for those time bins and $n_i = 0$ for all others. This strategy is called "time-stamping" because each photon arrival time is "stamped" into the memory of a computer somewhere. Correlation functions (to any order) can then be calculated later.

A more common strategy is to compute the correlation function electronically, using coincidence counting techniques, also known as "time-correlated photon counting." For example, the photodiode signal can be used to start a timer (a time-to-amplitude converter or time-to-digital converter), and the next signal used to stop the timer. The timer value is then recorded by a computer or multi-channel analyzer, and the process is repeated. A histogram of the time differences is proportional to $\left\langle n_i n_{i+\tau/\delta t} \right\rangle$, assuming 1) $n_i \leq 1$ and 2) $\langle n \rangle \tau/\delta t \ll 1$. This second restriction arises because the timer counts only until the first stop event. More sophisticated, "multi-stop" counters can circumvent this problem.

When two or more detectors are used and we count only pairs (or trios, quartets, etc.) of photons that arrive in the *same* time bin, we talk of "coincidence detection" and "coincidence counting." This gives a signal proportional to $\langle n_{A,i} n_{B,i} \rangle$ and can be implemented with very simple electronics, often nothing more than AND gates and inexpensive counters.

Chapter 6

Representations of quantum states of light

6.1 Introduction

So far, the states of the field we have considered, number states, vacuum, coherent states, squeezed states, are all *pure states*. In this section we develop several ways to describe *mixed states* in quantum optics. As in quantum mechanics, a mixed state is described by a *density operator*. Unlike most problems in quantum mechanics, we will find that although the *density matrix* exists, is not the most useful representation for many situations. We will thus develop representations of the density operator in terms of continuous degrees of freedom such as the quadratures X_1, X_2. These will be *phase space distributions*.

It turns out that there are many phase space distributions, and we will only be able to mention the most common ones. For a more complete treatment, we recommend the books by Scully and Zubairy, and by Walls and Milburn, and references therein.

6.2 Density operator

A mixed state is described by its density operator

$$\rho \equiv \sum w_i \left|\psi_i\right\rangle \left\langle\psi_i\right| \tag{6.1}$$

where $|\psi_i\rangle$ are normalized states and $w_i \geq 0$ and $\sum_i w_i = 1$. Thus $\{w_i\}$ can be interpreted as a probabilities: w_i is the probability that the system is prepared in the state $|\phi_i\rangle$. The expectation value of an operator A is

$$\langle A\rangle = \mathrm{Tr}[\rho A] \equiv \sum_j \left\langle\phi_j\right| \rho A \left|\phi_j\right\rangle \tag{6.2}$$

where $\{|\phi_j\rangle\}$ is a set of basis states. It follows that

$$\mathrm{Tr}[\rho A] = \sum_i w_i \langle\psi_i| A |\psi_i\rangle . \tag{6.3}$$

This describes an *incoherent* addition of the contributions from each $|\psi_i\rangle$.

6.3 Representation by number states

The density operator can be expanded in terms of number states as

$$\rho = \sum_{n,n'} \rho_{n,n'} |n\rangle \langle n'| \tag{6.4}$$

where *density matrix* is

$$\rho_{n,n'} \equiv \langle n| \rho |n'\rangle . \tag{6.5}$$

Note that this simple relationship is possible because

$$\sum_n |n\rangle \langle n| = I. \tag{6.6}$$

Not all expansions that we use will have this nice property.

This representation contains all the information about the state, and is simple to interpret. For example the diagonal element $\rho_{n,n}$ is the probability to have n photons in the state, while the off-diagonal element $\rho_{0,1}$ is the coherence between the $n = 0$ and $n = 1$ parts of the state.

This representation is useful for fields with a definite extent in space or in time. For example, for fields within a cavity (as in the Jaynes-Cummings model), or to characterize the total (i.e. integrated) field in a pulse. But there are many situations in which counting the number of photons is not natural, for example the field emitted by a continuous-wave laser. Also, while there are good techniques for measuring the diagonal elements (photon counting), it is not so easy to measure the off-diagonal elements. For these reasons, we need other representations.

6.4 Representation in terms of quadrature states

The density operator can be expanded in terms of quadrature states as

$$\rho = \int dX_1\, dX_1'\, |X_1\rangle \langle X_1| \rho |X_1'\rangle \langle X_1'| = \int dX_1\, dX_1'\, g(X_1, X_1') |X_1\rangle \langle X_1'| \tag{6.7}$$

where

$$g(X_1, X_1') \equiv \langle X_1| \rho |X_1'\rangle = \sum_i w_i \langle X_1|\psi_i\rangle \langle\psi_i|X_1'\rangle = \sum_i w_i \psi_i(X_1)\psi_i^*(X_1') \tag{6.8}$$

and $\psi_i(X_1) \equiv \langle X_1|\psi_i\rangle$. Clearly a similar expression could be written for expansion in X_2 or any generalized quadrature. This representation has the advantage of being closely connected to the wave-functions $\psi_i(X_1)$, and thus may be more intuitive than other representations. But in fact it is almost never used in quantum optics, because an equivalent representation, the Wigner distribution (described below), is more symmetric, is easier to interpret, and is easier to measure.

6.5 Representations in terms of coherent states

Consider an expansion of the density operator in coherent states

$$\rho = \int d^2\alpha\, d^2\alpha'\, f(\alpha,\alpha')\, |\alpha\rangle \left\langle \alpha' \right| \tag{6.9}$$

where $\alpha \equiv x_1 + ix_2 = r\exp[i\phi]$ and thus $d^2\alpha = dx_1 dx_2 = rdrd\phi$ [1]. The function f is analogous to the density matrix, and the expansion is always possible due to the over-completeness of the coherent states. I.e., there is always a function f which satisfies this. For example,

$$f(\alpha,\alpha') = \frac{1}{\pi^2} \langle\alpha|\, \rho \left|\alpha'\right\rangle \tag{6.10}$$

satisfies Equation (6.9), which is easily shown using the identity

$$\frac{1}{\pi}\int d^2\alpha\, |\alpha\rangle\langle\alpha| = I. \tag{6.11}$$

But this solution is not unique. For example, for the pure coherent state $\rho = |\beta\rangle\langle\beta|$, one solution is $f(\alpha,\alpha') = \langle\alpha|\beta\rangle\langle\beta|\alpha'\rangle/\pi^2 = \exp[-|\alpha-\beta|^2/2 + \alpha^*\beta]\exp[-|\alpha'-\beta|^2/2 + \alpha'\beta^*]/\pi^2$ and another solution is $f'(\alpha,\alpha') = \delta^2(\alpha-\beta)\delta^2(\alpha'-\beta)$. This suggests that this representation somehow has too many degrees of freedom. At the same time, we know from the expansion in quadrature states that the density operator can represented by a function of just two real variables, while the function f depends on four. This motivates us to look for lower-dimensional representations of the density operator.

6.5.1 Glauber-Sudarshan P-representation

If we assume that the f function above is *diagonal*, i.e. $f(\alpha,\alpha') = P(\alpha)\delta^2(\alpha-\alpha')$, then we have the expansion

$$\rho = \int d^2\alpha\, d^2\alpha'\, f(\alpha,\alpha')\, |\alpha\rangle \left\langle \alpha' \right| = \int d^2\alpha\, P(\alpha)\, |\alpha\rangle\langle\alpha|\,. \tag{6.12}$$

[1] This expansion is very similar to one considered by Glauber, namely $\rho = \frac{1}{\pi^2}\int d^2\alpha\, d^2\beta\, R(\alpha^*,\beta)\, |\alpha\rangle\langle\beta| \exp[-(|\alpha|^2+|\beta|^2)/2]$ where $R(\alpha^*,\beta) = \langle\alpha|\,\rho\,|\beta\rangle \exp[(|\alpha|^2+|\beta|^2)/2]$.

This representation was introduced independently by Glauber and Sudarshan, and is called the Glauber-Sudarshan P-representation or simply the P-representation. It can be shown that

$$\mathrm{Tr}[\rho] = 1 = \int d^2\alpha\, P(\alpha). \tag{6.13}$$

The function $P(\alpha)$ can sometimes be thought of as a probability distribution, and the state as a mixture of coherent states. This is possible when $P(\alpha) \geq 0$ for all α, but for some states this is not the case. For example, for squeezed states P is negative in some regions. For $n > 0$ number states, P does not exist, at least not as a regular function. But when it P does exist, it is uniquely determined by ρ.

6.5.2 Husimi distribution or Q-representation

Another representation of the state is

$$Q(\alpha) \equiv \frac{1}{\pi} \langle\alpha|\, \rho\, |\alpha\rangle\,. \tag{6.14}$$

Apart from a factor of π, this is the diagonal element of the function $f(\alpha, \alpha') = \langle\alpha|\, \rho\, |\alpha'\rangle\, /\pi^2$. It can be shown that

$$\int d^2\alpha\, Q(\alpha) = 1 \tag{6.15}$$

and clearly $Q(\alpha)$ is positive definite. Note that Q does not describe an expansion of the state, i.e., $\rho \neq \int d^2\alpha\, Q(\alpha)\, |\alpha\rangle\, \langle\alpha|$. Nevertheless, $Q(\alpha)$ determines uniquely the state ρ.

6.6 Wigner-Weyl distribution

The Wigner-Weyl distribution, also called the Wigner distribution and the Wigner function, is similar in many ways to the P- and Q-representations. Its shape in phase-space is somewhere between the two. Its mathematical definition is more complicated than the P- and Q-distributions', and because of this the Wigner function often seems rather mysterious. Nevertheless, it will be worth knowing because:

1) It exists for any state.
2) It corresponds to the classical phase-space distribution.
3) It has a Fourier-transform relationship to the density operator (in the quadrature representation).
4) It correctly predicts marginal distributions.
5) It can be measured (indirectly).

To introduce the Wigner function, we start with some classical statistics.

6.6.1 Classical phase-space distributions

In classical physics, an individual system follows a trajectory through phase space, defined by the evolution of the coordinates and momenta, e.g., $q(t), p(t)$. It is also possible to describe an ensemble of such systems behaving in a statistical manner, such that a function $F(q,p)$ describes the probability to find the system near to q,p, i.e., the probability to be in the range q to $q+dq$ and p to $p+dp$ is $F(q,p)dq\,dp$. This probability density F is a *phase-space distribution*. Some characteristics of such a distribution are: non-negativity $F \geq 0$, normalization $\int dq\,dp\,F(q,p) = 1$. The *marginal distributions* $F(q) \equiv \int dp\,F(q,p)$ and $F(p) \equiv \int dq\,F(q,p)$ give the probability density for a single coordinate or momentum, averaging over the possible values of the other degree of freedom. This generalizes in the obvious way to more coordinates and momenta.

For a classical harmonic oscillator, the direct method to measure $F(x,p)$ is by repeatedly preparing the state and measuring simultaneously x and p. After many measurements it is possible to estimate $F(x,p)$. There are also indirect methods, which do not require simultaneous measurement of x and p. One way to do this is by measuring the *characteristic function* for the state.

In classical statistics, a characteristic function $\chi(k)$ is defined as the expectation value of the random variable $\exp[ikx]$, where x itself is a random variable and k is a parameter. This can be calculated for any random variable x. For example, if x were the arrival time of your morning train, over the course of a year you could sample x 365 times, and then estimate $\langle\exp[ikx]\rangle \approx (\exp[ikx_1] + \exp[ikx_2] + ...)/365$. And of course, you can calculate this for any value of k you like.

If $F(x)$ is the distribution function (or probability density function) for x, then

$$\chi(k) = \left\langle e^{ikx} \right\rangle = \int dx\,F(x)e^{ikx}. \tag{6.16}$$

In multiple dimensions, this is generalized to

$$\chi(\mathbf{k}) = \left\langle e^{i\mathbf{k}\cdot\mathbf{x}} \right\rangle = \int d^n x\,F(\mathbf{x})e^{i\mathbf{k}\cdot\mathbf{x}}. \tag{6.17}$$

Thus the characteristic function is nothing more than the (inverse) Fourier transform of the distribution function. Note that for any given $\mathbf{k}$, $\chi(\mathbf{k})$ can be estimated by measurements of just one component of $\mathbf{x}$, the component parallel to $\mathbf{k}$. If this component can be directly measured, then no simultaneous measurements are required. The characteristic function is useful in statistics for the usual reasons, e.g., a convolution of distribution functions can be computed using the product of their characteristic functions.

6.6.2 Applying classical statistics to a quantum system

For a quantum system, direct measurement of the distribution function is not possible, because simultaneous measurement of X_1 and X_2 is forbidden by the uncertainty principle. The

characteristic function, in contrast, *can* be measured. For example, if we want to measure

$$\chi(\mathbf{k}) = \left\langle e^{i\mathbf{k}\cdot\mathbf{X}} \right\rangle \tag{6.18}$$

for some particular value of $\mathbf{k}$, where $\mathbf{k} \equiv (k_1, k_2) \equiv (k\cos\theta, k\sin\theta)$ and $\mathbf{X} = (X_1, X_2)$, then we have

$$\chi(\mathbf{k}) = \left\langle e^{ik(X_1\cos\theta + X_2\sin\theta)} \right\rangle = \left\langle e^{ikX_\theta} \right\rangle \tag{6.19}$$

where $X_\theta = X_1\cos\theta + X_2\sin\theta$ is a generalized quadrature. The value of $\chi(\mathbf{k})$ can be determined by measuring X_θ several times, for example in a balanced homodyne measurement, and then computing the average $\left\langle e^{ikX_\theta} \right\rangle$. This does not conflict with the uncertainty principle, because we only measure one quadrature at a time. By repeating this with several values of θ, it is possible to build up an approximation of $\chi(\mathbf{k})$. From quantum theory, we know that the expectation value is $\left\langle e^{ikX_\theta} \right\rangle = \mathrm{Tr}[\rho \exp[ikX_\theta]]$, so that

$$\chi(\mathbf{k}) = \mathrm{Tr}[\rho e^{i\mathbf{k}\cdot\mathbf{X}}]. \tag{6.20}$$

It is then possible to calculate a distribution function $W(x_1, x_2)$ as the Fourier transform of χ. Note that in $W(x_1, x_2)$, x_1, x_2 are real-valued parameters of the function W, not operators. Also, note that $\mathbf{k}\cdot\mathbf{X} = (k_1 + ik_2)a^\dagger + (k_1 - ik_2)a$. The distribution function is then

$$\begin{aligned} W(x_1, x_2) &= \mathcal{F}[\chi(k_1, k_2)] = \frac{1}{4\pi^2}\int dk_1\, dk_2\, e^{-i\mathbf{k}\cdot\mathbf{x}}\, \mathrm{Tr}\left[\rho e^{i\mathbf{k}\cdot\mathbf{X}}\right] \\ &= \frac{1}{4\pi^2}\int dk_1\, dk_2\, e^{-i\mathbf{k}\cdot\mathbf{x}}\, \mathrm{Tr}\left[\rho e^{i[(k_1+ik_2)a^\dagger + (k_1-ik_2)a]}\right]. \end{aligned} \tag{6.21}$$

We are almost finished. It only remains to put this in the usual form. We note that if $2\alpha = x_1 + ix_2$ and $\beta = k_1 + ik_2$, then $\alpha\beta^* + \alpha^*\beta = x_1k_1 + x_2k_2$ and the Fourier transform can be written equivalently

$$\begin{aligned} W(x_1, x_2) &= \mathcal{F}[\chi(k_1, k_2)] = \frac{1}{4\pi^2}\int dk_1\, dk_2\, e^{-ik_1x_1 - ik_2x_2}\chi(k_1, k_2) \\ W(\alpha) &= \mathcal{F}[\chi(\beta)] = \frac{1}{\pi^2}\int d^2\beta\, e^{-i(\alpha\beta^* + \alpha^*\beta)}\chi(\beta) \end{aligned} \tag{6.22}$$

where

$$\chi(\beta) = \mathrm{Tr}\left[\rho e^{i(\beta a^\dagger + \beta^* a)}\right]. \tag{6.23}$$

This agrees with the convention used by Scully and Zubairy. Other authors may define β differently, for example it is common to use $\beta' \equiv (-k_2 + ik_1)/2 = i\beta$, so that the Fourier transform is $\propto \int d^2\beta'\, \exp[\alpha\beta'^* - \alpha^*\beta']\chi'(\beta')$ and the characteristic function is $\chi'(\beta') = \left\langle \exp[\beta' a^\dagger - \beta'^* a] \right\rangle$. [2]. We have used the symbol $W(\alpha)$, because this in fact is our definition of the Wigner distribution: the Fourier transform of the characteristic function χ.

[2]The usual form for the Wigner function in quantum optics makes it look as much as possible like the P and Q distributions. For example the quadratures x_1, x_2 are organized into a single complex amplitude α as if they were a coherent state. This is elegant but perhaps misleading. There are no coherent states used in the definition of the Wigner function, and it is not related to expansion of ρ in coherent states. W is also more general than the P and Q distributions. You can define a Wigner function for other coordinate/momentum pairs, e.g. angle and angular momentum, for which coherent states do not exist and thus P and Q are not defined.

It is worth re-stating how we derived the Wigner distribution, to highlight what is classical and what is quantum about it. W is the Fourier transform of the characteristic function, which in classical statistics is the (inverse) Fourier transform of the distribution F, so classically we would have the trivial relationship $W = \mathcal{F}[\mathcal{F}^{-1}[F]] = F$. In Quantum mechanics we do not have a distribution F, but we have the density operator ρ, which predicts all measurements we can make, including χ, which plays the role of $\mathcal{F}^{-1}[F]$ (classically it *is* $\mathcal{F}^{-1}[F]$). The quantum ingredient is the measurement outcomes (and the way of predicting them), but the statistical treatment is completely classical.

6.6.3 Facts about the Wigner distribution

The Wigner distribution is normalized, in the sense that

$$\int W(x_1, x_2) dx_1 dx_2 = \frac{1}{4} W(\alpha) d^2\alpha = Tr[\rho]. \tag{6.24}$$

The Wigner distribution is real. The Wigner distribution can be used to calculate the overlap or fidelity of two states, in the sense that if W, W' are the Wigner distributions for states ρ, ρ', respectively, then

$$\int W(x_1, x_2) W'(x_1, x_2) dx_1 dx_2 = \frac{1}{4\pi} Tr[\rho\rho']. \tag{6.25}$$

Marginal distributions

For any ρ, measurements of the quadrature X_1 will have a probability distribution

$$P(x_1) = Tr[\rho \left|x_1\right\rangle \left\langle x_1\right|] \tag{6.26}$$

where $\left|x_1\right\rangle \left\langle x_1\right|$ is a projector onto the X_1 eigenstate $X_1 \left|x_1\right\rangle = x_1 \left|x_1\right\rangle$. As with position eigenstates, the eigenstates $\left|x_1\right\rangle$ are not normalized, but the projectors are, in a sense:

$$\int dx_1 \ \left|x_1\right\rangle \left\langle x_1\right| = I, \tag{6.27}$$

where I is the identity operator. More abstractly, we can write $\left|x_1\right\rangle \left\langle x_1\right| = \delta(X_1 - x_1)$.

We now calculate $P(x_1)$ as a marginal distribution. That is, we integrate W over the other coordinates (only x_2 in this case) and find

$$\begin{aligned}
\int dx_2 \, W(x_1, x_2) &= \frac{1}{4\pi^2} \int dx_2 \int dk_1 \, dk_2 \, e^{-ik_1x_1 - ik_2x_2} \chi(k_1, k_2) \\
&= \frac{1}{2\pi} \int dk_1 \, dk_2 \, \delta(k_2) e^{-ik_1x_1} \chi(k_1, k_2) \\
&= \frac{1}{2\pi} \int dk_1 \, e^{-ik_1x_1} \chi(k_1, 0) \\
&= \frac{1}{2\pi} \int dk_1 \, e^{-ik_1x_1} Tr[\rho e^{ik_1X_1}] \\
&= Tr[\rho\delta(X_1 - x_1)] = P(x_1).
\end{aligned} \tag{6.28}$$

This confirms our earlier claim, that W is like the probability distribution F; it can be used to calculate the probability for an observable simply by integrating over ("tracing over") the other observables.

To find the distribution for a generalized quadrature $X_\theta = X_1 \cos\theta + X_2 \sin\theta$, we define rotated coordinates $x_\theta, x_{\bar{\theta}}$ through

$$\begin{aligned} x_1 &= x_\theta \cos\theta - x_{\bar{\theta}} \sin\theta \\ x_2 &= x_\theta \sin\theta + x_{\bar{\theta}} \cos\theta \end{aligned} \tag{6.29}$$

so that

$$\begin{aligned} \int dx_{\bar{\theta}}\, W(x_1, x_2) &= \int dx_{\bar{\theta}}\, W(x_\theta \cos\theta - x_{\bar{\theta}} \sin\theta, x_\theta \sin\theta + x_{\bar{\theta}} \cos\theta) \\ &= Tr[\rho \delta(X_\theta - x_\theta)] = P_\theta(x_\theta). \end{aligned} \tag{6.30}$$

6.6.4 "Characteristic functions" for Q- and P-distributions

We can generalize the above mathematics (but not it's interpretation!) to include the P- and Q-distributions. In particular, we can identify three versions of the characteristic function. The one we defined above is called the *symmetrically-ordered characteristic function*

$$C(\beta) \equiv \left\langle e^{i(a\beta^* + a^\dagger \beta)} \right\rangle = \mathrm{Tr}[\rho e^{i(a\beta^* + a^\dagger \beta)}] = \chi(\beta). \tag{6.31}$$

The other two are the normally-ordered (N) and anti-normally-ordered (A) characteristic functions

$$C_N(\beta) \equiv \left\langle e^{ia^\dagger \beta} e^{ia\beta^*} \right\rangle = \mathrm{Tr}[\rho e^{ia^\dagger \beta} e^{ia\beta^*}] \tag{6.32}$$

and

$$C_A(\beta) \equiv \left\langle e^{ia\beta^*} e^{ia^\dagger \beta} \right\rangle = \mathrm{Tr}[\rho e^{ia\beta^*} e^{ia^\dagger \beta}]. \tag{6.33}$$

If we now take the Fourier transforms of these characteristic functions, we will have three distributions, which are equal to W, P, and Q.

$$W(\alpha) \equiv \frac{1}{\pi^2} \int d^2\beta\, e^{-i(\alpha\beta^* + \alpha^*\beta)} C(\beta) \tag{6.34}$$

$$P(\alpha) \equiv \frac{1}{\pi^2} \int d^2\beta\, e^{-i(\alpha\beta^* + \alpha^*\beta)} C_N(\beta) \tag{6.35}$$

$$Q(\alpha) \equiv \frac{1}{\pi^2} \int d^2\beta\, e^{-i(\alpha\beta^* + \alpha^*\beta)} C_A(\beta) \tag{6.36}$$

These characteristic functions are useful for proving things about the various distributions. Note that only the symmetrically-ordered version can be measured in the way we described above, and only W has an interpretation as a distribution of the quadratures X_1, X_2.

Chapter 7

Proofs of non-classicality

Many of the early experiments in quantum optics attempted to demonstrate differences between the quantum theory of light and classical theory. A very early example is Taylor's 1909 experiment, where a two-slit interference pattern was seen, even though the light was attenuated such that on average there was less than one photon in the apparatus at any time. This experiment failed to show any difference between quantum and classical optics, and in fact with our current understanding we do not expect any difference. As described above, the interference signal is a measure of the amplitude correlation function $G^{(1)}$. In contrast, experiments that measure $G^{(2)}$ do show clear differences between classical and quantum theory. Historically, this type of experiment served as a "proof" of the existence of photons.

7.1 Quantum vs. Classical (vs. Non-classical)

From the perspective of the philosophy of science, quantum optics and classical optics are both hypotheses about the behaviour of light. Because they disagree, at most one of these hypotheses can be true, and experiments can be done to disprove one or both hypotheses. This is similar to the situation of Newtonian mechanics vs. Einsteinian mechanics; precise measurements can distinguish between these two theories. And indeed, both Newtonian mechanics and classical optics were disproved in the first half of the 20th century.

There is an important difference, however. Newtonian and Einsteinian mechanics are both classical theories which give exact predictions about measurable quantities. In principle, their predictions differ by a measurable amount in any non-trivial situation. For example, in Newtonian gravitation, two orbiting bodies maintain a fixed average distance, while in Einsteinian gravitation they spiral inward toward each other. Testing the difference between these theories is a practical matter: can the measurement be made well enough? In contrast, many situations in optics give exactly the same predictions for both theories. An example is Taylor's 1909 experiment. Another way to view the question, many experimental observations can be explained by either theory, and thus do not distinguish between them.

This situation, in which the quantum theory and the classical one often agree, arises for basic reasons. The quantum theory was constructed to agree with the classical theory in many ways, most importantly in the average values of the fields. So any average field measurements will not distinguish them. The uncertainty principle, which places a lower limit on the fluctuations of the quantum fields, also does not immediately help in distinguishing the theories. For the classical fields, there is no lower limit to the fluctuations, but there is also no upper limit. Thus an observation of the fluctuations cannot disprove the classical theory. Observation of fields with fluctuations below the uncertainty principle limit would disprove quantum optics. But this has never been observed.

An important exercise is to distinguish between quantum states which can be used to disprove classical optics, and those which cannot. Closely tied to this is the distinction between observations that violate the classical assumptions and observations that are consistent with classical theory. Typically, it is possible to characterize classical predictions with an inequality relating some combination of correlation functions. When this inequality is violated, the classical theory is disproved. Also, the class of quantum states which can violate this inequality in experiment can be identified. These states are then considered *non-classical states*.[1]

7.2 $g^{(2)}(0)$

For example, if an optical field with intensity $I_0(t)$ is split on a beamsplitter and sent to two detectors, we measure two intensities $I_1(t) = \langle I_1 \rangle + \delta I_1(t)$ and $I_2(t) = \langle I_2 \rangle + \delta I_2(t)$.

At equal times $\tau = 0$,

$$g^{(2)}(0) = \frac{\langle I_1(t) I_2(t) \rangle}{\langle I_1 \rangle \langle I_2 \rangle} = 1 + \frac{\langle \delta I_1(t) \delta I_2(t) \rangle}{\langle I_1 \rangle \langle I_2 \rangle}. \tag{7.1}$$

In general, if $g^{(2)}(0) > 1$, there is a positive correlation between I_1 and I_2 and we say that the light is "bunched." If $g^{(2)}(0) < 1$ then I_1 and I_2 are anti-correlated and we say that the light is "anti-bunched."

Classically, we expect the beamsplitter to make a faithful copy, so that $I_1(t) = I_2(t) = I_0(t)$ and $\delta I_1 \delta I_2 \geq 0$. This means that

$$g^{(2)}(0) \geq 1 \quad \text{(classical result)}. \tag{7.2}$$

Quantum mechanics allows for $g^{(2)}(0)$ to be greater than or less than 1, even for $g^{(2)}(0)$ to be zero. For example, if the light field contains a definite number of photons, any photons

[1] Note that "non-classical states" are always described within the quantum theory, and are those which, in some ideal experiment, could be used to disprove classical optics. It is common to use the term "classical states" for those states, also described within the quantum theory, which are not "non-classical states." For example, it is common to say that coherent states are "classical states." This usage, while almost universally adopted, is still misleading. After all, all states within the quantum theory are quantum states, described by vectors in a Hilbert space (or mixtures of such states), as opposed to states within the classical theory, which would be described by vector fields, or probability distributions over the space of vector fields.

directed to one detector will not be detected by the other. This produces anti-correlations, i.e., anti-bunching. The simplest state with a definite number of photons is the single-photon state $|\phi\rangle_0 = |1\rangle_0 = \hat{a}_0^\dagger |0\rangle$. To treat this situation quantum mechanically, we note that due to the mixing at the beamsplitter $\hat{E}_1 = (\hat{E}_0 + \hat{E}_{empty})/\sqrt{2}$ and $\hat{E}_2 = (\hat{E}_0 - \hat{E}_{empty})/\sqrt{2}$. Here $\hat{E}_{empty}$ is the electric field operator for the mode that enters the beamsplitter by the unused port. Since we are not sending any light in here, this cannot cause detections, but we need to include it so that the action of the beamsplitter is unitary. We find that

$$G^{(2)}_{1,2}(\tau) = \frac{1}{2}\left\langle \hat{E}_0^{(-)}(t)\hat{E}_0^{(-)}(t+\tau)\hat{E}_0^{(+)}(t+\tau)\hat{E}_0^{(+)}(t)\right\rangle \tag{7.3}$$

plus terms containing $\hat{E}^{(+)}_{empty}$ or $\hat{E}^{(-)}_{empty}$ which give zero contribution. Since this contains two annihilation operators acting on the single-photon state $|1\rangle_0$, it gives zero. On the other hand, the average intensities are not zero:

$$\langle I_1\rangle = \left\langle \hat{E}_1^{(-)}(t)\hat{E}_1^{(+)}(t)\right\rangle = \frac{1}{2}\left\langle \hat{E}_0^{(-)}(t)\hat{E}_0^{(+)}(t)\right\rangle \neq 0 \tag{7.4}$$

with a similar expression for $\langle I_2\rangle$. As a result, the quantum prediction is

$$g^{(2)}(0) = \frac{G^{(2)}_{1,2}(0)}{\langle I_1\rangle \langle I_2\rangle} = 0 \quad \text{(1 photon state)}. \tag{7.5}$$

Note that to observe $g^{(2)}(0) < 1$ it is not necessary to have a deterministic single photon source, i.e., one which produces exactly the state $|1\rangle$. It is sufficient to have a stochastic source which sometimes produces $|1\rangle$, very rarely produces $|2\rangle, |3\rangle, ...$, and the rest of the time produces $|0\rangle$. For the pure state $|\psi\rangle = |0\rangle + \varepsilon|1\rangle$, or the mixed state $\rho = |0\rangle\langle 0| + \varepsilon^2 |1\rangle\langle 1|$, the same result $g^{(2)}(0) = 0$ would be observed.

7.2.1 Anti-bunching and the P-distribution

The function

$$g^{(2)}(0) \equiv \frac{\left\langle \hat{E}^{(-)}(t)\hat{E}^{(-)}(t)\hat{E}^{(+)}(t)\hat{E}^{(+)}(t)\right\rangle}{\left\langle \hat{E}^{(-)}(t)\hat{E}^{(+)}(t)\right\rangle^2} = 1 + \frac{\langle \delta I \delta I\rangle}{\langle I\rangle^2} \tag{7.6}$$

is a ratio of normally-ordered correlation functions, and for this reason can be simply calculated from the P-distribution, if it is known. Specifically,

$$g^{(2)}(0) = 1 + \frac{\int d^2\alpha\, P(\alpha)\left[\alpha^*\alpha - \langle \alpha^*\alpha\rangle\right]^2}{\left[\int d^2\alpha\, P(\alpha)\alpha^*\alpha\right]^2}. \tag{7.7}$$

The second term on the RHS contains all positive quantities, except for the distribution $P(\alpha)$ in the numerator. From this we can conclude that to observe $g^{(2)}(0) < 1$, the P-distribution for the state (if it exists) must be negative somewhere.

A similar statement can be made about squeezing. First we need a normally-ordered expression for the quadrature variance

$$\begin{aligned}\Delta X_1^2 &\equiv \left\langle \left[(a+a^\dagger) - \left\langle a+a^\dagger \right\rangle\right]^2 \right\rangle = \left\langle (a+a^\dagger)^2 - \left\langle a+a^\dagger \right\rangle^2 \right\rangle \\ &= \left\langle a^2 + aa^\dagger + a^\dagger a + (a^\dagger)^2 - \left\langle a+a^\dagger \right\rangle^2 \right\rangle \\ &= \left\langle a^2 + 1 + 2a^\dagger a + (a^\dagger)^2 - \left\langle a+a^\dagger \right\rangle^2 \right\rangle . \end{aligned} \tag{7.8}$$

We can calculate this expectation value with the P-distribution as

$$\begin{aligned}\Delta X_1^2 &\equiv 1 + \int d^2\alpha\, P(\alpha) \left[(\alpha+\alpha^*)^2 - \langle \alpha+\alpha^* \rangle^2\right] \\ &= 1 + \int d^2\alpha\, P(\alpha) \left[(\alpha+\alpha^*) - \langle \alpha+\alpha^* \rangle\right]^2 . \end{aligned} \tag{7.9}$$

Here too, we see that the condition $\Delta X_1^2 < 1$ implies negative values for the P-distribution. The same could be calculated for $\Delta X_2^2 < 1$.

Note that a negative P-distribution is necessary, but not sufficient, to give squeezing or anti-bunching.

7.3 $g^{(2)}(0)$ variant and the Cauchy-Schwarz inequality.

Sources which stochastically produce pairs of photons are particularly well suited to demonstrating non-classicality. Pairs of photons in different modes are produced nearly simultaneously, for example from a cascade transition in an atom or by parametric down-conversion. The state of the field might be

$$|\phi\rangle = \left(1 + \varepsilon \int dt'\, \hat{E}_{sig}^{(-)}(t') \hat{E}_{trig}^{(-)}(t') + O(\varepsilon^2)\right) |0\rangle . \tag{7.10}$$

Here $\varepsilon \ll 1$ so that most of the time the field contains no photons, but sometimes it contains one pair, and very rarely more than two photons. The integral over t' indicates that we do not know when the photons might have been produced. We will see later how such a state might arise.

One photon is detected by a "trigger" detector, which announces, or "heralds" the presence of the other photon. The signal photon is then sent to a beam-splitter and two detectors as above. This was the case in a famous experiment performed in 1986 by Grangier, et al. [2]. This three-detector coincidence experiment can be described as a measurement of the correlation function

$$G^{(3)}(0) = \left\langle \hat{E}_{sig}^{(-)}(t) \hat{E}_{sig}^{(-)}(t) \hat{E}_{trig}^{(-)}(t) \hat{E}_{trig}^{(+)}(t) \hat{E}_{sig}^{(+)}(t) \hat{E}_{sig}^{(+)}(t) \right\rangle_\phi . \tag{7.11}$$

[2] P. GRANGIER, G. ROGER, A. ASPECT, "Experimental evidence for a photon anticorrelation effect on a beam splitter: a new light on single - photon interferences". Europhysics Letters February 15th, 1986 pp 173-179.

But since the field at the trigger detector commutes with the others, we can push these operators to the outside, so that it is clear they act to annihilate one trigger photon.

$$\begin{aligned} G^{(3)}(0) &= \left\langle \hat{E}^{(-)}_{trig}(t)\hat{E}^{(-)}_{sig}(t)\hat{E}^{(-)}_{sig}(t)\hat{E}^{(+)}_{sig}(t)\hat{E}^{(+)}_{sig}(t)\hat{E}^{(+)}_{trig}(t)\right\rangle_{\phi} \\ &= \left\langle \hat{E}^{(-)}_{sig}(t)\hat{E}^{(-)}_{sig}(t)\hat{E}^{(+)}_{sig}(t)\hat{E}^{(+)}_{sig}(t)\right\rangle_{\phi'} \\ &= G^{(2)}_{\phi'}(0). \end{aligned} \tag{7.12}$$

Here $|\phi'\rangle \equiv \hat{E}^{(+)}_{trig}(t)\,|\phi\rangle$ is given by

$$\left|\phi'\right\rangle = \left(0 + \varepsilon\hat{E}^{(-)}_{sig}(t) + O(\varepsilon^2)\right)|0\rangle\,. \tag{7.13}$$

For small ε, this is effectively a single-photon state.

7.4 Cauchy-Schwarz inequality

Another way to show non-classical behaviour with photon pairs is by demonstrating a "violation of the Cauchy-Schwarz inequality."

For any vector space with an scalar product, for example the dot product $\vec{a}\cdot\vec{b}$ between vectors $\vec{a},\vec{b}$ in a Euclidean space, we can define a norm as $||\vec{a}|| \equiv \sqrt{\vec{a}\cdot\vec{a}} \geq 0$. The Cauchy-Schwarz inequality says

$$|\vec{a}\cdot\vec{b}| \leq ||\vec{a}||\;||\vec{b}||. \tag{7.14}$$

If we take $A(t), B(t)$ to be elements of the space of real-valued functions on the interval $0 \leq t \leq T$, then we can define the inner product

$$(A,B) \equiv \frac{1}{T}\int_0^T dt A(t)B(t) = \langle A(t)B(t)\rangle\,. \tag{7.15}$$

The Cauchy-Schwarz inequality for this situation then implies

$$\langle A(t)B(t)\rangle \leq \sqrt{\langle A^2(t)\rangle\,\langle B^2(t)\rangle} \tag{7.16}$$

or

$$|\,\langle A(t)B(t)\rangle\,|^2 \leq \left\langle A^2(t)\right\rangle\left\langle B^2(t)\right\rangle. \tag{7.17}$$

Much like in the test of $g^{(2)}$, the experiment uses beam-splitters to split the A and B modes to two detectors each $A \to A_1, A_2$ and $B \to B_1, B_2$. The photo-currents are correlated to find $\langle I_{A_1} I_{A_2}\rangle$, $\langle I_{B_1} I_{B_2}\rangle$, and $\langle (I_{A_1}+I_{A_2})(I_{B_1}+I_{B_2})\rangle$. Classically, we expect $I_{A_1} = I_{A_2} = I_A/2$ and $I_{B_1} = I_{B_2} = I_B/2$, so that the Cauchy-Schwarz inequality implies

$$|\,\langle (I_{A_1}+I_{A_2})(I_{B_1}+I_{B_2})\rangle\,|^2 \leq 16\,\langle I_{A_1} I_{A_2}\rangle\,\langle I_{B_1} I_{B_2}\rangle\,. \quad \text{(classical result)} \tag{7.18}$$

In words, two classical intensities must be better correlated with themselves than they are with each other.

Quantum mechanically, we expect $\langle I_{A_1} I_{A_2} \rangle = \left\langle \hat{E}_A^{(-)} \hat{E}_A^{(-)} \hat{E}_A^{(+)} \hat{E}_A^{(+)} \right\rangle /4$ and $\langle I_{B_1} I_{B_2} \rangle = \left\langle \hat{E}_B^{(-)} \hat{E}_B^{(-)} \hat{E}_B^{(+)} \hat{E}_B^{(+)} \right\rangle /4$ and $\langle (I_{A_1} + I_{A_2})(I_{B_1} + I_{B_2}) \rangle = \left\langle \hat{E}_A^{(-)} \hat{E}_B^{(-)} \hat{E}_B^{(+)} \hat{E}_A^{(+)} \right\rangle$. In the case where the state is $|\phi\rangle = \hat{a}_A^\dagger \hat{a}_B^\dagger |0\rangle$,

$$\langle I_{A_1} I_{A_2} \rangle = \langle I_{B_1} I_{B_2} \rangle = 0 \quad \text{(2 photon result)} \tag{7.19}$$

and

$$\langle (I_{A_1} + I_{A_2})(I_{B_1} + I_{B_2}) \rangle > 0 \quad \text{(2 photon result)} \tag{7.20}$$

so that the classical result above (often called the Cauchy-Schwarz inequality) is violated.

7.5 Bell inequalities

In 1935, Albert Einstein, along with two colleagues, published an attack on quantum theory that would become known the "Einstein-Podolsky-Rosen Paradox." In the paper, they claimed to demonstrate that quantum mechanics was incomplete, in the sense that it should be possible to make exact, deterministic predictions about the outcomes of experiments, not just probabilistic predictions. In 1964, Bell showed by means of an inequality that any theory that contained such exact predictions must also contain action-at-a-distance, i.e., non-locality. Bell's inequality was, at least in principle, testable by experiment. Clauser, Horne, Shimony and Holt introduced more convenient inequalities showing the same thing, and the first experiments began around 1974, with contradictory results, some supporting quantum theory and some supporting the alternative, "local hidden variable theories." In 1983 and 1984, the experiments of Aspect and colleagues showed very clearly support for quantum theory and the disproving of all local hidden variable theories, subject to some reasonable assumptions regarding loopholes in the experimental proof. At the time of writing this, work continues to close the remaining loopholes and there are some exciting proposals for decisive experiments in the area. For a fuller discussion, see D. Dehlinger and M. W. Mitchell, "Entangled photons, nonlocality, and Bell inequalities in the undergraduate laboratory," Am. J. Phys. 70, 903-910 (2002).

7.6 Squeezing

Squeezing is obviously important in quantum optics, and often is taken to indicate non-classical behaviour. But the relation between squeezing and the testing of classical vs. quantum optics is not so simple as for photon-counting experiments. This is because, while classical optics describes precisely the behaviour of the field, it does not completely specify the model for *detection* of fields. In other words, a model for detectors is needed to complete the calculation and arrive to a prediction for how much noise will be observed.

7.6.1 Classical noise in the fields

In classical optics, there is no uncertainty principle to force the field to contain noise. A classical field can have exact, noise-free values. So the *existence* of a low-noise field (squeezed below the uncertainty principle limit, for example), is completely consistent with classical optics. Nevertheless, the *observation*, i.e., the detection, of a low-noise field may or may not be possible according to classical optics, depending on the model of detection.

7.6.2 Classical square-law detector

Direct detection, which converts light power into a signal, requires a "square-law" detector (the output signal is proportional to the square of the input field). A classical version of this would be an antenna connected to a resistor, with a thermometer to measure the rise in temperature of the resistor. The voltage produced in the antenna is proportional to the incident field, and the power generated in the resistor is proportional to the voltage squared.

Within classical physics, this kind of detector can be noiseless: the only classical noise source is thermal noise in the resistor, and the resistor could in principle be cooled arbitrarily close to zero temperature.

7.6.3 Semi-classical square-law detector

The classical square-law detector, based on thermal effects, is not very similar to real optical detectors, which typically involve the excitation of electrons (of a semiconductor in a photodiode, or from a metal in a photo-multiplier tube, or from an atom in a Geiger counter). We understand that these are quantum systems, and thus we expect them to behave quantum mechanically. We can build a semi-classical model of detection, in which the quantum detector interacts with a classical radiation field. In this case, the detector is described by a Hamiltonian

$$\begin{aligned} H_{det} &\equiv \sum_i H_i \\ H_i &\equiv H_i^{(0)} + \sum_i \mathbf{d}_i \cdot \mathbf{E}(\mathbf{x}_i, t), \end{aligned} \tag{7.21}$$

where $H_i^{(0)}$ is the internal Hamiltonian of the atom which contains the i'th electron, $\mathbf{x}_i$ and $\mathbf{d}_i$ are its position and dipole moment operator (this can cause transitions) and $\mathbf{E}$ is the classical electric field.

If the effect of the field is to excite an electron from a definite ground state $|g\rangle$ into a continuum of possible excited states $|f\rangle$, then the rate of excitation is given by Fermi's "golden rule"

$$w_{g\to f} = \frac{2\pi}{\hbar} |\langle f| \mathbf{d}_i \cdot \tilde{\mathbf{E}}(\mathbf{x}_i, \omega_{fg}) |g\rangle|^2 \rho \tag{7.22}$$

where $\tilde{\mathbf{E}}(\mathbf{x}_i, \omega_{fg})$ is the field component at the transition frequency ω_{fg} and ρ is the density of final states. This clearly is proportional to $|\mathbf{E}|^2$, so this is a square-law detector.

Each electron in the detector will be excited with this rate, and each electron behaves independently of the others. This is a consequence of the Hamiltonian we assumed, which does not include interactions among electrons.[3] The total number of excited electrons will be random, and follow a Poisson distribution. If the average number of electrons that make an upward transition in a time δt is

$$\langle N_{fg} \rangle = \sum_i w_{g \to f} \delta t, \tag{7.23}$$

then the RMS deviation in N_{fg} is

$$\delta N_{fg} = \sqrt{\langle N_{fg} \rangle}. \tag{7.24}$$

This model contains "shot noise," but it does not come from the photons, but rather from the model of the detector. Since each electron sees the same field, and each electron responds in a probabilistic manner and independently of the others, the result is a probabilistic (noisy) response. Note that, for this semi-classical model, any direct detection signal will have shot noise. But as we have seen in Chapter 4, our other kind of detection (homodyne) is built from direct detection, so this will also have shot noise.

7.6.4 Fully quantum detection

The Glauber theory of photo-detection uses the same model of light interacting with the detector, but the field **E** is an operator. This can produce strong correlations in the electrons of the detector. For example, when a single-photon state illuminates a large-area detector, each part of the detector interacts with the field. But unlike the semi-classical case, only one part of the detector can be excited, because the field only contains one photon. This is a very strong form of correlation: if electron i is excited, then all electrons $j \neq i$ will not be.

A quantum state which contains on average $\langle N \rangle$ photons per unit time and a fluctuation of $\delta N < \langle N \rangle$ in the same time, has sub-Poissonian power fluctuations. This is called "intensity squeezed" light. In the Glauber theory, when this light is used to illuminate a high-efficiency detector, the detected signal will also be sub-Poissonian. Thus a fully-quantum theory can predict detection of intensity squeezing, which the semi-classical theory cannot. The same is true of quadrature squeezing.

7.6.5 Anti-bunching and the P-distribution

The function

$$g^{(2)}(0) \equiv \frac{\left\langle \hat{E}^{(-)}(t)\hat{E}^{(-)}(t)\hat{E}^{(+)}(t)\hat{E}^{(+)}(t) \right\rangle}{\left\langle \hat{E}^{(-)}(t)\hat{E}^{(+)}(t) \right\rangle^2} = 1 + \frac{\langle \delta I \delta I \rangle}{\langle I \rangle^2} \tag{7.25}$$

[3] This is an additional assumption that would have to be justified for the particular type of detector. For example in a Geiger counter the electrons are each from different atoms in a gas, and thus not in contact. In general, any detector that is large enough will at least have *regions* that behave independently.

is a ratio of normally-ordered correlation functions, and for this reason can be simply calculated from the P-distribution, if it is known. Specifically,

$$g^{(2)}(0) = 1 + \frac{\int d^2\alpha\, P(\alpha)\left[\alpha^*\alpha - \langle\alpha^*\alpha\rangle\right]^2}{\left[\int d^2\alpha\, P(\alpha)\alpha^*\alpha\right]^2}. \tag{7.26}$$

The second term on the RHS contains all positive quantities, except for the distribution $P(\alpha)$ in the numerator. From this we can conclude that to observe $g^{(2)}(0) < 1$, the P-distribution for the state (if it exists) must be negative somewhere.

A similar statement can be made about squeezing. First we need a normally-ordered expression for the quadrature variance

$$\begin{aligned}
\Delta X_1^2 &\equiv \left\langle \left[(a + a^\dagger) - \left\langle a + a^\dagger \right\rangle\right]^2 \right\rangle = \left\langle (a + a^\dagger)^2 - \left\langle a + a^\dagger \right\rangle^2 \right\rangle \\
&= \left\langle a^2 + aa^\dagger + a^\dagger a + (a^\dagger)^2 - \left\langle a + a^\dagger \right\rangle^2 \right\rangle \\
&= \left\langle a^2 + 1 + 2a^\dagger a + (a^\dagger)^2 - \left\langle a + a^\dagger \right\rangle^2 \right\rangle.
\end{aligned} \tag{7.27}$$

We can calculate this expectation value with the P-distribution as

$$\begin{aligned}
\Delta X_1^2 &\equiv 1 + \int d^2\alpha\, P(\alpha)\left[(\alpha + \alpha^*)^2 - \langle\alpha + \alpha^*\rangle^2\right] \\
&= 1 + \int d^2\alpha\, P(\alpha)\left[(\alpha + \alpha^*) - \langle\alpha + \alpha^*\rangle\right]^2.
\end{aligned} \tag{7.28}$$

Here too, we see that the condition $\Delta X_1^2 < 1$ implies negative values for the P-distribution. The same could be calculated for $\Delta X_2^2 < 1$.

Note that a negative P-distribution is necessary, but not sufficient, to give squeezing or anti-bunching.

Chapter 8

Behaviour of quantum fields in linear optics

In this chapter we consider the behaviour of quantum fields in linear optical systems. A "system" here is very general, so general in fact that it is difficult to give a helpful definition, but we try anyway: A linear optical system transforms its inputs A, which are fields in some space-time region, into outputs $f(A)$, which are fields in some (possibly different) space-time region, by a transformation which is linear: $f(A+B) = f(A) + f(B)$.

Almost everything in optics before the laser is linear in this sense. A few examples of linear optical systems (and the linear effects they use): a prism (refraction), a grating (diffraction), a lens (refraction again), a beamsplitter (partial reflection), an interferometer (interference), a neutral-density filter (absorption), a laser amplifier (linear amplification). Do we need a quantum theory for each of these things? No, thank goodness! Quantum fields in linear optical systems behave very much like classical fields, and almost everything in a classic book like Born and Wolf's "Principles of Optics" applies equally well to quantum optics as to classical optics.

There are some differences, however, in the area of losses and amplification. Related to this we mention squeezing, which is sometimes considered to be linear (it fits the definition above) but more often is considered to be non-linear optics, because all laboratory squeezers have used non-linear processes to generate the squeezing (more about that in the next chapter).

8.1 Diffraction

Diffraction is the first "system" we encounter in optics; it describes the change in an optical field as it passes through empty space. We present here the simplest theory of diffraction, the diffraction of a scalar field. This is a good approximation for a single polarization in the paraxial regime. Traditionally, this subject has been treated using Huygens' principle, where the field in a space can be found by considering the field on the boundary of the space as radiating sources. This gives rise to Franhofer and Kirchoff diffraction integrals. A more modern approach, better

adapted to optical beams, is to use the paraxial wave equation to propagate the field forward in space.

Fraunhofer and Kirchoff diffraction

If we know a field $\hat{E}(\mathbf{r}, t)$ on an aperture, then diffraction theory gives the field

$$\hat{E}(\mathbf{r}', t') \propto \int d^2\mathbf{r}\, \frac{\hat{E}(\mathbf{r}, t' - |\mathbf{r} - \mathbf{r}'|/c)}{|\mathbf{r} - \mathbf{r}'|} \tag{8.1}$$

in the volume. If we consider just one frequency component, we find that $\hat{E}(\mathbf{r}, t' - \tau) = \hat{E}^{(+)}(\mathbf{r}, t') \exp[i\omega\tau] + \hat{E}^{(-)}(\mathbf{r}, t') \exp[-i\omega\tau]$, and we make the usual approximations to get the Fraunhofer diffraction integral

$$\hat{E}^{(+)}(\mathbf{r}') \propto \int d^2\mathbf{r}\, \hat{E}^{(+)}(\mathbf{r}) e^{-ik\mathbf{r}'\cdot\mathbf{r}/R} \tag{8.2}$$

where $k = \omega/c$ is the wavenumber and R is the distance from the aperture.

example: beating the diffraction limit

As an example, suppose that we produce pairs of photons behind the two slits of a double-slit apparatus, with each pair emerging from the same slit. A state with this property would be $|\phi\rangle = \frac{1}{2}(\hat{a}^\dagger(0)\hat{a}^\dagger(0) + \hat{a}^\dagger(d)\hat{a}^\dagger(d))\,|0\rangle$, where d is the separation of the slits. Does this state show diffraction? At the slits, the coherence is

$$\left\langle \hat{E}^{(-)}(x_1)\hat{E}^{(+)}(x_2) \right\rangle \propto [\delta(x_1) + \delta(x_1 - d)]\delta(x_1 - x_2) \tag{8.3}$$

so that in the far field

$$\begin{aligned} P_1(x') &\propto \left\langle \hat{E}^{(-)}(x')\hat{E}^{(+)}(x') \right\rangle \\ &= \int dx_1 dx_2 e^{ikx(x_1 - x_2)/R} \left\langle \hat{E}^{(-)}(x_1)\hat{E}^{(+)}(x_2) \right\rangle \\ &= const. \end{aligned} \tag{8.4}$$

Thus there is no one-photon diffraction pattern.

There is, however, a two-photon diffraction pattern. The relevant correlation function is

$$\begin{aligned} \left\langle \hat{E}^{(-)}(x_1)\hat{E}^{(-)}(x_2)\hat{E}^{(+)}(x_3)\hat{E}^{(+)}(x_4) \right\rangle \propto\ & [\delta(x_1) + \delta(x_1 - d)]\delta(x_1 - x_2) \\ & \times[\delta(x_3) + \delta(x_3 - d)]\delta(x_3 - x_4). \end{aligned} \tag{8.5}$$

The factors $\delta(x_1 - x_2)$ and $\delta(x_3 - x_4)$ arise because the two photons pass through the same slit. In the far-field, the probability of seeing two photons arrive at the same position x is

$$\begin{aligned} P_2(x') &\propto \left\langle \hat{E}^{(-)}(x')\hat{E}^{(-)}(x')\hat{E}^{(+)}(x')\hat{E}^{(+)}(x') \right\rangle \\ &= \int dx_1 dx_2 dx_3 dx_4 e^{ikx(x_1+x_2-x_3-x_4)/R} \\ &\quad \times \left\langle \hat{E}^{(-)}(x_1)\hat{E}^{(-)}(x_2)\hat{E}^{(+)}(x_3)\hat{E}^{(+)}(x_4) \right\rangle . \end{aligned} \tag{8.6}$$

By virtue of the delta functions, this evaluates to

$$P_2(x) \propto 2e^{ikx(0)/R} + e^{i2kxd/R} + e^{-i2kxd/R} = 2[1 + \cos(2kxd/R)]. \tag{8.7}$$

Remarkably, this interference pattern is finer by a factor of two than the ordinary interference from a double slit of width d. In principle, this can give spatial resolution better than the diffraction limit.

8.2 Paraxial wave equation

With the invention of lasers (and computers), another way of treating diffraction problems has become popular, and in fact is very useful for quantum optics. This is based on the paraxial wave equation (described in detail in Appendix **??**)

$$\left[\nabla_T^2 + 2ik(\partial_z + \frac{n}{c}\partial_t)\right] \mathcal{E}^{(+)}(\mathbf{x}, t) = 0 \tag{8.8}$$

where $\mathcal{E}^{(+)}(\mathbf{x}, t)$ is the envelope of the positive-frequency part of the field, such that the field itself is

$$E(\mathbf{x}, t) = \mathcal{E}^{(+)}(\mathbf{x}, t)e^{ikz - i\omega t} + H.c. \tag{8.9}$$

Here $\exp[ikz - i\omega t]$ is the carrier wave of the field, and by assumption $\mathcal{E}^{(+)}(\mathbf{x}, t)$ is slowly-varying in both position and time.

8.3 Linear optical elements

Most optical elements: mirrors, lenses, beam-splitters, wave-plates, etc. are both linear and (approximately) lossless. Generically, they produce transformations on the optical field $E_{out}(r, t) = f[E_{in}(r, t)]$ where f is linear and invertible. The diffraction integral above is an example. For linear optics, the behaviour of quantum fields is exactly the same as for the equivalent classical fields, i.e., the equation holds whether the Es are classical fields or quantum field operators.

8.3.1 beam-splitter

The lowly beam splitter is at the heart of many stunning quantum optics experiments. It can transform product states into entangled states, quadrature squeezed states into Einstein-Podolsky-Rosen states and produce quantum logic gates. We have already seen that it is used for quadrature detection and multi-photon detection in measuring quantum correlations.

A beam splitter has four ports, call them A, B, C, D, assuming that on transmission $A \to C, B \to D$ and on reflection $A \to D, B \to C$. It is natural to introduce two coordinate systems S_A, S_B each with the centre of the beam-splitter at the origin, and related by a reflection about the beam-splitter surface. The fields $\{\hat{E}_A(\mathbf{x}_A, t), \hat{E}_C(\mathbf{x}_C, t)\}$ and $\{\hat{E}_B(\mathbf{x}_B, t), \hat{E}_D(\mathbf{x}_D, t)\}$ are considered in coordinate system S_A and S_B, respectively.

The action of the beam-splitter is the unitary transformation

$$\begin{pmatrix} E_C^{(+)}(\mathbf{x}_C, t) \\ E_D^{(+)}(\mathbf{x}_D, t) \end{pmatrix} = U \begin{pmatrix} E_A^{(+)}(\mathbf{x}_A, t) \\ E_B^{(+)}(\mathbf{x}_B, t) \end{pmatrix} \tag{8.10}$$

where U is a unitary matrix which contains the amplitudes for transmission and reflection. Given a mode decomposition of the fields, this means that the creation and annihilation operators transform as

$$\begin{pmatrix} a_{k,C} \\ a_{k,D} \end{pmatrix} = U \begin{pmatrix} a_{k,A} \\ a_{k,B} \end{pmatrix} \tag{8.11}$$

Omitting a global phase, a general form for a unitary matrix U is

$$U = e^{i\psi\sigma_z} e^{i\theta\sigma_y} e^{i\phi\sigma_z} = \begin{pmatrix} e^{i\psi} & 0 \\ 0 & e^{-i\psi} \end{pmatrix} \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} e^{i\phi} & 0 \\ 0 & e^{-i\phi} \end{pmatrix} \tag{8.12}$$

where the σs are the Pauli matrices and θ, ϕ, ψ are known as Euler angles. In practice, the mixing angle θ is easy to measure, and the phases ϕ, ψ are very difficult, as they are equivalent to the phases along the paths to/from the beam-splitter. (To measure them with an interferometer one would have to know the precise distance in each of the paths A, B, C, D, for example)[1]. For this reason, ϕ, ψ are usually chosen for calculational convenience. For 50/50 beam splitters common choices are

$$U = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \text{ or } \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & i \\ i & 1 \end{pmatrix}. \tag{8.13}$$

Example: Hong-Ou-Mandel effect ("single-mode" version)

Suppose we have an input state that has one photon at each input A, B, and the photons are in matched modes k (i.e., mirror image modes that have the same spatio-temporal shape).

[1] It is, however, possible to measure *relative* phases, for example between two polarizations, in a single beam splitter. For example, a beam-splitter might transform 45° polarized light at A into right circularly-polarized light at D. Thus $\phi_H - \phi_V$ would be known to be $\pi/2$.

The mode k might be single-frequency mode with gaussian mode shape, or a single-photon wave-packet, the only requirement is that it is the same on each side. We write the state

$$|\phi\rangle = a_A^\dagger a_B^\dagger |0\rangle , \tag{8.14}$$

where the mode index k is suppressed, since it will be the same for A, B, C, D.

We can look at the intensities at the outputs of the beam-splitter by first calculating the correlation functions $\left\langle a_C^\dagger a_C \right\rangle, \left\langle a_D^\dagger a_D \right\rangle$. ϕ, ψ can be taken as zero with no loss of generality. We find

$$\begin{aligned} a_C^\dagger a_C &= (\cos\theta a_A^\dagger - \sin\theta a_B^\dagger)(\cos\theta a_A - \sin\theta a_B) \\ &= \cos^2\theta a_A^\dagger a_A + \sin^2\theta a_B^\dagger a_B - \sin\theta\cos\theta(a_A^\dagger a_B + a_B^\dagger a_A) \end{aligned} \tag{8.15}$$

and

$$a_D^\dagger a_D = \sin^2\theta a_A^\dagger a_A + \cos^2\theta a_B^\dagger a_B - \sin\theta\cos\theta(a_A^\dagger a_B + a_B^\dagger a_A) \tag{8.16}$$

from which $\left\langle a_C^\dagger a_C \right\rangle = \left\langle a_D^\dagger a_D \right\rangle = 1$, independent of θ, as well as ϕ, ψ. Thus there is no interference observed by looking at intensities.

On the other hand, if we look at the coincidence probability $\left\langle a_C^\dagger a_D^\dagger a_D a_C \right\rangle$, we find

$$\begin{aligned} a_D a_C &= (\sin\theta a_A + \cos\theta a_B)(\cos\theta a_A - \sin\theta a_B) \\ &= (\cos^2\theta - \sin^2\theta) a_A a_B + O(a_A^2, a_B^2) \end{aligned} \tag{8.17}$$

so that

$$\left\langle a_C^\dagger a_D^\dagger a_D a_C \right\rangle = \cos^2 2\theta. \tag{8.18}$$

Thus there *is* interference in coincidence detection indeed there is perfect interference visibility. This is referred to as multi-photon interference, higher-order interference[2] or non-classical interference.

We can also understand this by looking at the state, written in terms of C, D operators. Note that we are still in the Heisenberg picture, and the state does not *change*, we are just re-writing it in a way that its properties at the output are evident. We assume $\phi = \psi = 0$, such that $a_A = \cos\theta a_C + \sin\theta a_D$ and $a_B = -\sin\theta a_C + \cos\theta a_D$. The product

$$\begin{aligned} a_A a_B &= (\cos\theta a_C + \sin\theta a_D)(-\sin\theta a_C + \cos\theta a_D) \\ &= \cos\theta\sin\theta[a_D^2 - a_C^2] + (\cos^2\theta - \sin^2\theta) a_C a_D. \end{aligned} \tag{8.19}$$

The state is then

$$|\phi\rangle = \sin 2\theta[\frac{1}{2}(a_D^\dagger)^2 - \frac{1}{2}(a_C^\dagger)^2]\,|0\rangle + \cos 2\theta a_C^\dagger a_D^\dagger\,|0\rangle \tag{8.20}$$

[2] Confusingly, the HOM effect is sometimes described as "second-order interference" and sometimes as "fourth-order interference." The confusion comes from $G^{(2)}$, which contains two intensities, or equivalently four fields.

or in terms of photon numbers $|n_C, n_D\rangle$,

$$|\phi\rangle = \sin 2\theta \frac{1}{\sqrt{2}}(|0,2\rangle - |2,0\rangle) + \cos 2\theta \, |1,1\rangle \tag{8.21}$$

8.4 Loss and Gain

For very general reasons, loss in an optical system is always accompanied by noise, or extra fluctuations associated with the losses. We can illustrate this with a model of a lossy system, which is simply a beam-splitter that removes some of the field we are interested in. If the input mode is A, and the output mode is C, with nothing (vacuum) input into modes B and D, we have as before

$$\begin{pmatrix} E_C^{(+)}(\mathbf{x}_C, t) \\ E_D^{(+)}(\mathbf{x}_D, t) \end{pmatrix} = U \begin{pmatrix} E_A^{(+)}(\mathbf{x}_A, t) \\ E_B^{(+)}(\mathbf{x}_B, t) \end{pmatrix}, \tag{8.22}$$

and we assume U has the form

$$U = \begin{pmatrix} t & r \\ -r & t \end{pmatrix} \tag{8.23}$$

with t and r real and $|t|^2 + |r|^2 = 1$. We see that the output field is

$$E_C^{(+)}(\mathbf{x}_C, t) = tE_A^{(+)}(\mathbf{x}_A, t) + rE_B^{(+)}(\mathbf{x}_B, t). \tag{8.24}$$

In terms of annihilation operators, this is simply

$$a_C = ta_A + ra_B \tag{8.25}$$

In classical optics we would be able to ignore the second term (because there is no light input to port B), but in quantum optics this extra field operator is a source of noise. For example, if we compute the variance of the quadrature $X_{1,C} = a_C + a_C^\dagger$, we find

$$\begin{aligned} \left\langle X_{1,C}^2 \right\rangle - \langle X_{1,C} \rangle^2 &= \left\langle t^2 X_{1,A}^2 \right\rangle + 2 \langle rt X_{1,A} X_{1,B} \rangle + \left\langle r^2 X_{1,B}^2 \right\rangle - (t \langle X_{1,A} \rangle + r \langle X_{1,B} \rangle)^2 \\ &= t^2 \left\langle X_{1,A}^2 - \langle X_{1,A} \rangle^2 \right\rangle + r^2 \left\langle X_{1,B}^2 \right\rangle \end{aligned} \tag{8.26}$$

where we have used the fact that $[X_{1,A}, X_{1,B}] = 0$ and $\langle X_{1,B} \rangle = 0$, since the input state on port B is vacuum. If we define $\eta \equiv t^2$ as the efficiency of transmission, we have

$$\mathrm{var}(X_{1,C}) = \eta \mathrm{var}(X_{1,A}) + (1-\eta)\mathrm{var}(X_{1,B}), \tag{8.27}$$

which is usually summarized by saying that any losses introduce $1 - \eta$ "units" of vacuum noise. In practice, it is often assumed that other losses (from scattering, material absorption, misalignment, etc.) also follow this rule, and often this is true. As we will see below, a careful analysis shows that this rule only specifies the lower limit on the noise caused by losses.

8.4.1 linear amplifiers and attenuators

The beam-splitter model gives a very useful result, and shows clearly (at least in this situation) where the noise comes from, and why it is unavoidable. In one of the most important papers in quantum optics, Carleton Caves showed a more general version of this, which applies both to amplifiers and to lossy processes. It also showed that phase-sensitive amplifiers (such as squeezers) could avoid quantum noise that affects phase-insensitive amplifiers (such as laser amplifiers).[3]

Assume a process has inputs $a, a^\dagger$ and outputs $b, b^\dagger$, with a linear relationship between the two

$$b = Ma + La^\dagger + F. \tag{8.28}$$

Here M and L are c-numbers, and F is a "noise operator." For the moment, all we know about this operator is that it does not depend on a or $a^\dagger$. We assume the process preserves the commutation relation

$$[b, b^\dagger] \quad = \quad [a, a^\dagger] = 1, \tag{8.29}$$

from which we immediately find

$$[F, F^\dagger] = 1 - |M|^2 + |L|^2. \tag{8.30}$$

Since this commutator is a real constant, F is something like an annihilation operator (if the RHS is positive) or like a creation operator (if the RHS is negative). It also implies an uncertainty relation. If we define $X_F \equiv F + F^\dagger$ and $P_F \equiv i(F^\dagger - F)$ we find that $[X_F, P_F] = 2i(1 - |M|^2 + |L|^2)$, which implies

$$\delta X_F \delta P_F \geq \left|1 - |M|^2 + |L|^2\right| \tag{8.31}$$

8.4.2 phase-insensitive case

We now assume that the amplifier is phase-insensitive. This imposes a condition on L, M and also one on F. If we make the phase rotation $a \rightarrow \exp[i\phi]a$ (so that $a^\dagger \rightarrow \exp[-i\phi]a^\dagger$), the output power is (ignoring the noise operator)

$$b^\dagger b = |M|^2 a^\dagger a + LM(a^2 e^{2i\phi} + (a^\dagger)^2 e^{-2i\phi}) + |L|^2 a a^\dagger. \tag{8.32}$$

For this to be phase invariant, we must have $LM = 0$. We are thus left with two possibilities:

$$b = Ma + F \tag{8.33}$$

which is a phase-preserving amplifier, and

$$b = La^\dagger + F \tag{8.34}$$

[3] C. M. Caves, "Quantum limits on noise in linear amplifiers," Phys. Rev. D, **26** 1817 (1982)

which is a phase-conjugating amplifier. We will only consider the phase-preserving case. We note that the power gain of the amplifier is $G \equiv |M|^2$. We also assume the noise operator is phase-insensitive, in the sense that

$$\left\langle f(F, F^\dagger) \right\rangle = \left\langle f(Fe^{i\phi}, F^\dagger e^{-i\phi}) \right\rangle \tag{8.35}$$

for any function f. For example, $\langle F \rangle$ must equal $\langle F \exp[i\pi] \rangle$, which is only possible if $\langle F \rangle$ is zero. Nevertheless, $\left\langle FF^\dagger \right\rangle$ could be anything. This also means that the uncertainty is equally distributed, $\mathrm{var}(\cos\theta X_F + \sin\theta P_F) = \mathrm{var}(X_F)$. Using the independence of F and a, for example $\langle aF \rangle = \langle a \rangle \langle F \rangle = 0$, we calculate the output fluctuations

$$\begin{aligned} \mathrm{var}(b + b^\dagger) &= \left\langle (Ma + F + M^* a^\dagger + F^\dagger)^2 \right\rangle - \left\langle Ma + F + M^* a^\dagger + F^\dagger \right\rangle^2 \\ &= \left\langle (Ma + M^* a^\dagger)^2 \right\rangle + \left\langle (F + F^\dagger)^2 \right\rangle - \left\langle Ma + M^* a^\dagger \right\rangle^2 - \left\langle F + F^\dagger \right\rangle^2 \\ &= \mathrm{var}(Ma + M^* a^\dagger) + \mathrm{var}(X_F). \end{aligned} \tag{8.36}$$

The first term in this expression is clearly the amplified (or attenuated) input fluctuation, while the second term is the noise added by the amplifier. Because of the uncertainty relation

$$\mathrm{var}(X_F) \geq \left| 1 - |M|^2 \right| = |1 - G| \,. \tag{8.37}$$

Of course, since the amplifier is phase-insensitive, if we calculate the noise in the other quadrature, we get the same result, namely

$$\mathrm{var}(ib^\dagger - ib) = \mathrm{var}(iM^* a^\dagger - iMa) + \mathrm{var}(P_F) \tag{8.38}$$

and

$$\mathrm{var}(P_F) \geq \left| 1 - |M|^2 \right| = |1 - G| \,. \tag{8.39}$$

If the gain is less than one, this lower noise limit agrees with the beam-splitter result above, with $\eta = G$. I.e., the beam-splitter introduces the minimum possible noise for a given (phase-insensitive) attenuation. In the case of an amplifier, the usual practice is to "refer the noise to the input," i.e., to model the amplifier as 1) the addition of noise F_{in} to the signal, followed by 2) noiseless amplification of signal + F_{in}. For this to give an output noise $\mathrm{var}(X_F)$, we must have $G\mathrm{var}(X_{F,in}) = \mathrm{var}(X_F)$, so that

$$\mathrm{var}(X_{F,in}) = \left| 1 - G^{-1} \right| . \tag{8.40}$$

In the case of large gain G, this becomes $\mathrm{var}(X_{F,in}) \to 1$. Recalling that for vacuum $\mathrm{var}(X) = 1$, we see that the noise added by a high-gain amplifier is the same as if vacuum noise were added at the input, and then amplified. This noise is in addition to any noise on the input. For example, if the input is the state $|n = 0\rangle$, i.e., vacuum, with $\mathrm{var}(X^{(\mathrm{in})}) = \mathrm{var}(P^{(\mathrm{in})}) = 1$, and the power gain is $G \gg 1$, then the output has $\mathrm{var}(X^{(\mathrm{out})}) = G(\mathrm{var}(X^{(\mathrm{in})}) + 1 - G^{-1}) \approx 2G$.

8.4.3 phase-sensitive amplifiers

If we return to Equation 8.30, we can easily see the conditions for noiseless amplification, i.e. for $[F, F^\dagger] = 0$, namely $|M|^2 - |L|^2 = 1$. If we write the general solution as $M = \exp[i\psi]\cosh r$, $L = \exp[i\psi]\exp[-2i\phi]\sinh r$, we have

$$\begin{aligned} b &= e^{i\psi}\left[a\cosh r + a^\dagger e^{-2i\phi}\sinh r\right] \\ &= e^{i\psi}\left[S^\dagger(\varepsilon)\hat{a}S(\varepsilon)\right] \end{aligned} \tag{8.41}$$

where $S(\varepsilon)$ is the squeeze operator of Chapter 3 (see Equation 3.21). We see that, apart from a phase shift, the output operator b is the squeezed input $S^\dagger(\varepsilon)\hat{a}S(\varepsilon)$. This means that squeezing is (in principle) noiseless. But to get this result we assumed very little, simply that the process was linear. It seems that the only linear, noiseless amplifiers are squeezers!

Chapter 9

Quantum Sensing and Metrology

Quantum systems are extraordinarily good at measurement, much better than classical systems could ever be. One reason has to do with energy quantization. In atoms, this creates discrete levels, and allows the system to be initialized in a definite state. In contrast, in classical physics all states are part of a continuum - it is always possible to add or remove a bit of energy. This, together with thermal physics, means that classical systems are always noisy. This is maybe easiest to illustrate with light. Consider blackbody radiation, which would contribute a noise to any optical measurement, including vision using your own eyes.

We can use the Bose-Einstein distribution to compute the average energy in a field mode of angular frequency ω:

$$\langle H_{\mathrm{EM}} \rangle \quad = \quad \begin{cases} \hbar\omega\langle n\rangle = \hbar\omega(e^{\hbar\omega/k_B T}-1)^{-1} & \text{quantum} \\ k_B T & \text{classical} \end{cases} \tag{9.1}$$

The classical result follows from the equipartition theorem, and approximates the quantum result when $\hbar\omega \ll k_B T$, which describes the Rayleigh-Jeans regime. If we consider $T = 300\,\mathrm{K}$, so that $k_B T/\hbar \approx 2\pi \times 6 \times 10^{12}\,\mathrm{Hz}$ and visible light with $\omega \sim 2\pi \times 6 \times 10^{14}\,\mathrm{Hz}$, we find that the classical energy per mode corresponds to $\approx 1 \times 10^{-2}$ photons. This looks small, but the number of modes is huge. In a single-mode fibre, for example, we would have to consider all of the frequency modes, covering a bandwidth of roughly $1 \times 10^{15}\,\mathrm{Hz}$, giving about 1×10^{13} photons/s worth of optical power due to (classical) blackbody radiation. This noise source would be a real problem for any optical sensing. Fortunately the quantum result, which is a consequence of energy quantization, gives $\langle n\rangle \sim 1 \times 10^{-42}$. Even multiplied by the many modes, the blackbody contribution is an utterly irrelevant 1×10^{-27} photons/s.

So we should be very happy that we live in a quantum world, where it is possible to see and to make optical measurements without being flooded with blackbody radiation. Quantum physics brings its own intrinsic noise sources. These are due to uncertainty relations, and a large part of quantum sensing is understanding and dealing with the consequences of these uncertainty relations on measurement. It turns out that if you understand them, you can make better measurements.

9.0.1 Metrology and Sensing

A measurement ultimately aims at an accurate estimate of a quantity, with both a small error and an estimate of the likely size of that error. The art and science of making good measurements is known as *metrology*. Metrology is a large field encompassing many related activities, including

- systems of units, e.g. the SI units
- development and use of *standards*, which are agreed-upon ways to measure in a system of units. For example, a cesium fountain clock is the primary standard for the atomic definition of the second.
- many technologies for specific measurements
- practices for estimating and reporting the systematic and statistical uncertainties on measured quantities

Quantum physics contributes to metrology in a number of ways. Starting in the mid-20th century, there has been a gradual shift from standards based on artefacts, e.g. the standard meter or the standard kilogram, which were pieces of metal kept under controlled conditions, to atomic standards. In 1960 the meter was re-defined in terms of the wavelength of a transition in krypton 86. In 1968 the second was re-defined in terms of the hyperfine transition frequency of atomic cesium. In 2018 a change was adopted making all the SI units traceable to fundamental constants like the speed of light and Planck's constant.

When units are defined in terms of quantum physical quantities, the standards used to measure them, e.g. atomic clocks, are quantum physical instruments. Quantum optics makes some interesting contributions here. For example, the Coulomb is since 2018 defined in terms of the charge of the electron, and devices for delivering individual electrons at a known rate, i.e. a flow of perfectly anti-bunched electrons, can be used as current standards. Similarly, non-classical photon sources are studied as possible standards for the Candela, a unit of optical power.

Many of the most precise measurements are interferometric in nature. This includes the already mentioned measurements of length using optical interferometers, but also atomic instruments are interferometers, in the sense that they take a quantum system, split it into a superposition of states, allow the states to acquire different phases through some physical mechanism (e.g. the passage of time, acceleration, or fields), and then interfere different the parts of the superposition to detect the effect of the phase shift. This describes atomic clocks (for measuring time), atom interferometers (for measuring acceleration, gravity and rotation), and atomic magnetometers (for measuring magnetic and radio-frequency fields).

Colloquially, we often speak of the accuracy and precision of a measurement. The precision is a measure of the statistical error in repeated measurements, whereas the accuracy is a measure of the error that would remain if you were able to take many measurements and average them, in order to make negligible all statistical errors. These concepts are closely related to the concepts of statistical and systematic errors when reporting measurement results. For example: An optical

spectrum analyzer might have a precision of 1 GHz and an accuracy of 10 GHz. If you measure the same, constant-frequency laser with this spectrum analyzer, repeated measurements will have a scatter of ~ 1 GHz about a value that itself might be different from the true value by 10 GHz. In this example the precision is better than the accuracy.

Quantum optics distinguishes itself from classical optics because it predicts different statistical properties for optical fields. This naturally implies different statistics for measurements based on optical fields. It is clear that this will have an effect on measurement precision, in the sense of statistical uncertainties. The intelligent use of quantum statistics, for example the use of squeezed states, can reduce statistical uncertainties. Quantum statistical effects can in a few cases also have a direct effect on measurement accuracy, for example the proposed use of non-classical light to define optical power standards.

Sensing describes the measurement of physical quantities arising in the environment of a sensor. For example, a pressure sensor measures (senses) a pressure, or a voltage sensor measures a voltage. Whereas metrology is concerned primarily with reproducible measurement of stable quantities, sensing is often concerned with measurement of transient or one-time signals. For example, a gravitational-wave detector gets only one chance to measure a gravitational wave as it passes through the Earth. Sensing is a problem in which statistical effects are of paramount importance. This is both because there is no opportunity to average multiple measurements to improve the precision, and because accuracy is far less important when measuring transient signals. A systematic error that produces a constant offset in the signal does not prevent us from measuring a transient on top of that offset.

In sensing we usually use the term "sensitivity" to describe the statistical uncertainty of a measurement, rather than precision, although it is the same concept. Note that all the terms "accuracy," "precision," and "sensitivity" have a very annoying ambiguity. The term is used to indicate the size of the error or uncertainty, as when we said that the optical spectrum analyzer had a precision of 1 GHz. In this sense, a small precision indicates a better measurement (less statistical uncertainty). Yet when we describe a measurement as "high precision," we mean it is a good measurement. So "high precision" means the precision is small. Crazy but true.

Improving the precision of optical interferometry, specifically for gravitational wave detection, is one of the original impulses for the development of quantum optics.

9.1 Estimation theory

From the perspective of statistics, any measurement is a problem in parameter estimation. Consider a random variable X with a probability distribution $P_\theta(X)$, where θ is a *parameter* of the PDF. This parameter is what we wish to learn about, but it is only accessible through measurements of X. We imagine making M measurements of X, to obtain the data $\vec{x} = (x_1, \ldots, x_M)$, sampled from the distribution $P_\theta(X)$. From these data, possibly supplemented by some prior information about θ, we will estimate θ.

An *estimator* is a function that takes as input the data $\vec{x}$, and gives as output an *estimate* for θ.

We will use a hat to indicate both the estimate $\hat{\theta}$ and the estimator $\hat{\theta}(\vec{x})$, when there is no risk of confusion. The estimator, as a function of random variables, is itself a random variable. For different purposes, some estimators are better suited than others. Important properties include

- The bias of $\hat{\theta}$ is defined as $E[\hat{\theta}] - \theta$, where θ is the true value of the parameter. Obviously, we would prefer that the estimator gets the right answer on average, so this is a basic requirement. A biased estimator contributes to systematic errors and reduces the accuracy of the measurement.

- The mean squared error (MSE) of the estimate is $E[(\hat{\theta} - \theta)^2] = \text{var}\hat{\theta} + E[\hat{\theta}]^2$. The MSE gives a measure of the total error, including both imprecision (in the variance term) and the bias. Other things equal, we would prefer an estimator with smaller MSE.

9.1.1 naive estimator

A naive estimator would calculate the sample mean $\mu_S \equiv N^{-1} \sum_{i=1}^{N} x_i$ (this is the only reasonable estimator for the mean) and then compute $\hat{\theta}_{\text{naive}} = \mu^{-1}(\mu_S)$, where μ^{-1} is the inverse of $\mu(\theta)$. This naive estimate is usually not the best estimate. We note for example

$$E[\hat{\theta}_{\text{naive}}] = E[\mu^{-1}(\mu_S)] \tag{9.2}$$

is equal to θ if $\mu(\theta)$ is linear, and thus μ^{-1} is also linear. But in other cases the naive estimator will be biased.

9.1.2 maximum likelihood estimator

Another common choice is the *maximum likelihood* (ML) estimator (MLE). We first define the likelihood of the data

$$\mathcal{L}(\mu) \equiv P_\theta(\vec{X} = \vec{x}) = \prod_{i=1}^{N} P_\theta(X = x_i) \tag{9.3}$$

where the last equality holds if the individual x_i are independent. The MLE is then defined as the value of θ that maximizes $\mathcal{L}$

$$\hat{\theta}_{\text{ML}} = \max_\theta \mathcal{L} = \max_\theta \ln \mathcal{L} \tag{9.4}$$

where the last form involves the log-likelihood $\ln \mathcal{L}$, which is usually easier to calculate with.

9.1.3 Fisher information

The Fisher information is a measure of how well a parameter can be estimated from a given distribution.

$$\begin{aligned} \mathcal{I}(\theta) &\equiv E[(\partial_\theta \ln P_\theta(X))^2] && (9.5)\\ &= E\left[(\frac{\partial_\theta P_\theta(X)}{P_\theta(X)})^2\right] && (9.6)\\ &= E[-\partial_\theta^2 \ln P_\theta(X)] && (9.7) \end{aligned}$$

where the last line follows only if $\partial_\theta^2 \ln P_\theta(X)$ exists. Note that the expectation implies an average over the values that X can take on, for example

$$E[(\partial_\theta \ln P_\theta(X))^2] \equiv \sum_{x\in\mathcal{X}} P_\theta(X=x)(\partial_\theta \ln P_\theta(X=x))^2. \quad (9.8)$$

$\mathcal{I}(\theta)$ is thus not a function of X. In contrast, it usually will depend on θ. If so, it means that the measurement's precision will not be equal for all values of the measured parameter.

If the the parameter θ has units $[\theta]$, then Fisher information has units $[1/\theta^2]$. This can be seen from the fact that ∂_θ has units $[1/\theta]$, while everything else inside the expectation $E[\cdot]$ is unitless.

The Fisher information of independent measurements is additive: if you measure X_1, distributed as $P_\theta^{(1)}(X_1)$ and this gives you Fisher information $\mathcal{I}_1(\theta)$, then you measure X_2, distributed as $P_\theta^{(2)}(X_2)$ and this gives you Fisher information $\mathcal{I}_2(\theta)$, this can be viewed as a single joint measurement of (X_1, X_2) distributed as $P_\theta^{(1,2)}(X_1, X_2) = P_\theta^{(1)}(X_1)P_\theta^{(2)}(X_2)$. The Fisher information for this joint measurement is $\mathcal{I}_{1,2}(\theta) = \mathcal{I}_1(\theta) + \mathcal{I}_2(\theta)$. It immediately follows that independently measuring the same variable M times, the total Fisher information is M times the Fisher information in a single measurement.

9.1.4 Cramer-Rao bound

The Fisher information is important for estimation because, for an unbiased estimator $\hat{\theta}$, the variance is lower bounded by the inverse of the Fisher information:

$$\text{var}(\hat{\theta}) \geq \frac{1}{\mathcal{I}(\theta)}. \quad (9.9)$$

This is then a limit on how well we can estimate θ by measurements of X. Although no unbiased estimator can beat this limit[1], it is possible for an unbiased estimator to do worse than this. So we still have the question of what estimator to use.

[1]Note that it is easy to beat this limit with a *biased* estimator. We could for example use the estimator $\hat{\theta}_0 = 0$, which always estimates that θ is zero, independently of the data. This is a biased estimator, because $E[\hat{\theta}_0] \neq \theta$, and in fact it is completely useless, since it does not depend in any way on the thing we are trying to measure, but it does have zero variance.

9.1.5 asymptotic case

Remembering that $\hat{\theta}$ is a random variable, we can ask about its distribution. This will of course depend on the estimator in question, but in general, we expect that with more data, the distribution of $\hat{\theta}$ will become narrower. We might even think that it will approach a normal distribution, based on our intuitions from the central limit theorem. This is not true for all estimators, but it is true for the MLE. It can be proven that as the M, the number of measurements, becomes large, $\hat{\theta}_{\mathrm{ML}}$ approaches a normal distribution. This is referred to as the asymptotic regime. What is more, in that same limit, the MLE saturates the Cramer-Rao bound, i.e.,

$$\mathrm{var}(\hat{\theta}_{\mathrm{ML}}) \xrightarrow[M\to\infty]{} \frac{1}{M\mathcal{I}(\theta)}. \tag{9.10}$$

9.1.6 resume

To summarize, if we have a random variable X and know its probability distribution $P_\theta(X)$, and we plan to collect a large number M of independent samples of X, then the variance of the MLE (provided the distribution is not pathological) will be

$$\mathrm{var}(\hat{\theta}_{\mathrm{ML}}) \xrightarrow[M\to\infty]{} \frac{1}{M\mathcal{I}(\theta)}, \tag{9.11}$$

where

$$\mathcal{I}(\theta) \equiv E[(\partial_\theta \ln P_\theta(X))^2] \tag{9.12}$$

is the Fisher information for a single measurement of X.

9.2 Application to quantum measurements

The above-described estimation theory can be applied to any scenario in which probability distributions are known. Quantum physics is of course an excellent resource for deriving probability distributions, and so it is very natural to apply this estimation theory to quantum problems. We remind ourselves of quantum measurements. Any measurement is described by a set of positive operators $\{O_i\}$, which must sum to the identity. This describes a positive operator-valued measure (POVM). We use the random variable X to represent the outcome of the measurement, which gives value x_i. The Born rule states that the probability of this outcome is

$$P(X = x_i) = \mathrm{Tr}[O_i\rho] \tag{9.13}$$

This is not yet a parametrized distribution, but it easily can become one. All we need is to make ρ a function of some parameter, and write [2].

$$P_\theta(X = x_i) = \mathrm{Tr}[O_i\rho(\theta)]. \tag{9.14}$$

[2] We could equally well use parametrized measurement operators $O_i(\theta)$, but in the spirit of quantum measurements and quantum sensing, we will assume that we know the O_i, which describe our laboratory equipment, but are uncertain about ρ.

For example, if the state is some fiducial state ρ_0 subjected to a unitary transformation U_θ, e.g. a rotation by an angle θ, the

$$\rho(\theta) = U_\theta \rho_0 U_\theta^\dagger \tag{9.15}$$

9.2.1 example: spin-1/2 system

We give a paradigmatic example. A spin-1/2 atom has a Hilbert space spanned by the states $|\pm 1/2\rangle$, indicating the spin projection along the z direction. We use this basis to describe state vectors and operators: $(+1/2, -1/2)^T$. In the presence of a magnetic field B along the z direction, the spin evolves under the Hamiltonian

$$H = \hbar\omega_L s_z = \hbar\omega_L \begin{pmatrix} 1/2 & 0 \\ 0 & -1/2 \end{pmatrix} \tag{9.16}$$

to produce the time-evolution operator

$$U(t) = e^{-iHt/\hbar} = \begin{pmatrix} \exp[-i\phi/2] & 0 \\ 0 & \exp[-i\phi/2] \end{pmatrix} \tag{9.17}$$

where $\phi = \omega_L t = \gamma B t$ is the precession angle. For any given initial state ρ_0, the state at time t is

$$\rho(t) = U(t)\rho_0 U^\dagger(t) \tag{9.18}$$

In particular an initial superposition state $|\psi(0)\rangle = (|+1/2\rangle + |-1/2\rangle)/\sqrt{2}$, so that $\rho_0 = |\psi(0)\rangle\langle\psi(0)| = \frac{1}{2}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$ the time-evolved density matrix is

$$\rho(t) = \frac{1}{2}\begin{pmatrix} 1 & \exp[-i\phi/2] \\ \exp[i\phi/2] & 1 \end{pmatrix}. \tag{9.19}$$

Different measurements are of course possible, but we consider a Stern-Gerlach measurement that projects onto the states $|\pm x\rangle \equiv (|1/2\rangle \pm |-1/2\rangle)/\sqrt{2}$. We call the result the random variable $X \in \{1/2, -1/2\}$, and the operators are projectors

$$O_{\pm 1/2} = |\pm x\rangle\langle\pm x| = \frac{1}{2}\begin{pmatrix} 1 & \pm 1 \\ \pm 1 & 1 \end{pmatrix}. \tag{9.20}$$

Then, if ϕ is the parameter we want to measure, we have

$$P_\phi(X = +1/2) = \mathrm{Tr}[O_{+1/2}\rho(\phi)] = \cos^2\frac{\phi}{2} = \frac{1}{2}(1 + \cos\phi) \tag{9.21}$$

$$P_\theta(X = -1/2) = \mathrm{Tr}[O_{-1/2}\rho(\phi)] = \sin^2\frac{\phi}{2} = \frac{1}{2}(1 - \cos\phi) \tag{9.22}$$

and thus

$$\partial_\phi P_\phi(X = +1/2) = -\sin\frac{\phi}{2}\cos\frac{\phi}{2} = -\frac{1}{2}\sin\phi \tag{9.23}$$

$$\partial_\phi P_\phi(X = -1/2) = \sin\frac{\phi}{2}\cos\frac{\phi}{2} = \frac{1}{2}\sin\phi \tag{9.24}$$

Inserting these into Equation 11.6, can now calculate the Fisher information

$$\begin{aligned}\mathcal{I}(\phi) &= \sum_{x\in\{+1/2,-1/2\}} P_\phi(X=x)\frac{(\partial_\phi P_\phi(X=x))^2}{(P_\phi(X=x))^2} \\ &= \frac{(-\frac{1}{2}\sin\phi)^2}{\frac{1}{2}(1+\cos\phi)} + \frac{(\frac{1}{2}\sin\phi)^2}{\frac{1}{2}(1-\cos\phi)} = 1 \end{aligned} \tag{9.25}$$

Remarkably, the Fisher information in this case is constant, independent of ϕ. If we use a single spin M times to measure a rotation angle, the Cramer-Rao bound tells us the variance on any unbiased estimator will be

$$\text{var}(\hat{\phi}) \geq \frac{1}{M} \tag{9.26}$$

for any value of the true angle ϕ. We also know that the MLE saturates this bound (in the limit of large M). We now find the MLE.

Suppose we make M measurements and observe data $\vec{x} = \{x_1, \ldots x_M\}$. The probability of seeing this (the likelihood function) is

$$\mathcal{L} = \prod_{i=1}^{N} P_\phi(X=x_i) = \frac{1}{2^M}(1+\cos\phi)^{m_+}(1-\cos\phi)^{m_-} \tag{9.27}$$

where $m_\pm$ is the number of $\{x_i\}$ equal to $\pm 1/2$. The log-likelihood is

$$\ln\mathcal{L} = \sum_{i=1}^{N} P_\phi(X=x_i) \propto m_+\ln(1+\cos\phi) + m_-\ln(1-\cos\phi) \tag{9.28}$$

which is maximized when

$$\partial_\phi \ln\mathcal{L} \propto -\frac{m_+\sin\phi}{1+\cos\phi} + \frac{m_-\sin\phi}{1-\cos\phi} = 0 \tag{9.29}$$

or

$$\frac{1+\cos\phi}{1-\cos\phi} = \frac{m_+}{m_-} \tag{9.30}$$

which is solved by

$$\hat{\phi} = \arccos\frac{m_+ - m_-}{m_+ + m_-}. \tag{9.31}$$

Chapter 10

Quantum fields in nonlinear optics

Although a great many situations can be treated with linear optics, as described in the previous chapter, nonlinear optics plays an essential role in quantum optics. Nonlinear optics studies optical phenomena in which one light field interacts with another, typically through the nonlinear susceptibility of a medium. The field is of great practical importance, and important classical applications include frequency conversion, all-optical modulation and soliton generation. In quantum optics, non-linear processes almost always produce some kind of non-classical state, and have been proposed as ways to create interactions between different quantum fields, for example using cross-phase modulation in which one beam causes a phase shift of another beam. If extended to the single-photon level, this would allow non-demolition measurement of photon number and quantum logic with photons.

10.1 Linear and nonlinear optics

The starting point for any treatment of optics is the Maxwell equations (ME), which we write here in their macroscopic form

$$\nabla \cdot \mathbf{D} = 0 \tag{10.1}$$

$$\nabla \cdot \mathbf{B} = 0 \tag{10.2}$$

$$\nabla \times \mathbf{E} = -\mu_0 \frac{\partial \mathbf{H}}{\partial t} \tag{10.3}$$

$$\nabla \times \mathbf{H} = \frac{\partial \mathbf{D}}{\partial t} \tag{10.4}$$

Where $\mathbf{D} \equiv \varepsilon_0 \mathbf{E} + \mathbf{P}$ and $\mathbf{H} \equiv \mathbf{B}/\mu_0 - \mathbf{M}$. This is a set of partial differential equations which is linear in the fields $\mathbf{E}, \mathbf{B}$ if the polarizations $\mathbf{P}, \mathbf{M}$ are linear functions of the fields $\mathbf{E}, \mathbf{B}$. In this situation a superposition principle holds, i.e., if $\mathbf{E}_1(\mathbf{x}, t), \mathbf{B}_1(\mathbf{x}, t)$ and $\mathbf{E}_2(\mathbf{x}, t), \mathbf{B}_2(\mathbf{x}, t)$ are solutions, then the sum of these is also a solution.

Linear equations are much easier to solve than nonlinear ones, but they are also very boring. For example, we consider the effect of these equations on the creation and annihilation operators in

the fields **E**, **B**. Since the fields are Hermitian, the ME can be reduced to the equation

$$\partial_t(c_i a_i + c_i^* a_i^\dagger) = M_{ij}(c_j a_j + c_j^* a_j^\dagger) \tag{10.5}$$

where i is an index for the modes and M (which we don't need to specify) is a real matrix that contains all the information about mode structure as well as the constants ε_0, μ_0. Because the equation is linear, we can write the solution as

$$\begin{aligned} a_i(t) &= L_{ij}(t)a_j(0) \\ a_i^\dagger(t) &= L_{ij}(t)a_j^\dagger(0), \end{aligned} \tag{10.6}$$

where $L_{ij} = \exp[tM_{ij}]$. We now consider the effect of this evolution on the quantum state of light. If the initial state $|\psi\rangle$ is a multi-mode coherent state, defined by

$$a_i(0)\,|\psi\rangle = \alpha_i\,|\psi\rangle \tag{10.7}$$

then at time t the operators a have evolved such that

$$a_i(t)\,|\psi\rangle = L_{ij}(t)a_j(0)\,|\psi\rangle = L_{ij}(t)\alpha_j\,|\psi\rangle\,. \tag{10.8}$$

But this also fits the definition of a multi-mode coherent state. If we had used the Schrödinger picture, we would have found that the state $|\psi\rangle$ evolves, but always remains a coherent state. It seems that linear optics, with coherent state inputs, always gives coherent state outputs. This is boring!

On the other hand, if we started with a nonlinear equation such as

$$\partial_t(c_i a_i + c_i^* a_i^\dagger) = M_{ij}(c_j a_j + c_j^* a_j^\dagger)^2 \tag{10.9}$$

then $a(t)$ would contain terms of the form $a^2(0)$, $a(0)a^\dagger(0)$, etc. From coherent state inputs we could get other sorts of outputs.

Because of this, all sources of interesting quantum states of light use nonlinear optics. This includes classic nonlinear optical materials such as crystals, but also clouds of atoms and even single ions and atoms, which are, in some ways, the most nonlinear of all.

10.2 Phenomenological approach

We start with one of the most versatile approaches, which we call "phenomenological," because it does not concern itself with the microscopic composition of the material, only its optical properties as expressed in the susceptibilities $\chi^{(1)}$, $\chi^{(2)}$, etc. This is a good technique for working with transparent linear and nonlinear materials, such as nonlinear crystals. It is not good for working with absorptive materials, or materials that have memory, such as atoms driven near resonance.

We assume that the polarization of the material obeys the usual expansion from nonlinear optics

$$P_i = \varepsilon_0[\chi_{ij}^{(1)}E_j + \chi_{ijk}^{(2)}E_jE_k + \chi_{ijkl}^{(3)}E_jE_kE_l + ...]. \tag{10.10}$$

To reduce clutter, we will not write the tensor indices unless they are necessary. In general, the susceptibilities χ are functions of several frequencies and can be complex. Fortunately the most common situations involve transparent materials and beams with bandwidths small compared to the bandgap of the crystal. These mean that χ is approximately real and approximately frequency-independent. Within these approximations, we can identify two main calculational approaches.

Effective Hamiltonian

Before we work out what the effective Hamiltonian is, we guess the result. We know that linear optics will produce oscillators with the dispersion relation $\omega_k = ck/n = ck/\sqrt{1+\chi}$. So we guess the Hamiltonian

$$\hat{H} = \hat{H}_0 + \hat{H}' \tag{10.11}$$

$$\hat{H}_0 = \frac{\varepsilon}{2}\int d^3r \hat{E}^2 + \frac{1}{2\mu_0}\int d^3r \hat{B}^2 = \sum_k \hbar\omega_k(\hat{a}_k^\dagger \hat{a}_k + 1/2) \tag{10.12}$$

where $\hat{H}'$ is the nonlinear contribution. We know that $\chi^{(2)}$ processes will include sum- and difference-frequency generation, described by normal-ordered terms like $a^\dagger_{\omega_1+\omega_2}a_{\omega_1}a_{\omega_2}$ and its Hermitian conjugate. This term appears in the expansion of $\hat{E}^3$. $\chi^{(3)}$ processes will contain terms of the form $a^\dagger a^\dagger aa$ and $a^\dagger aaa$ and their conjugates, which appear in $\hat{E}^4$. This suggests that there should be a parts of the Hamiltonian proportional to $\chi^{(2)}\hat{E}^3$, $\chi^{(3)}\hat{E}^4$, etc. We guess the following:

$$\hat{H}' = \int d^3r \left[C_2 : \hat{E}^3 : + C_3 : \hat{E}^4 : + ...\right] \tag{10.13}$$

where C_n are constants related to the various orders of χ and : : indicates normal-ordering. This guess is correct if $C_2 = -\varepsilon_0\chi^{(2)}/3$ and $C_3 = (\chi^{(2)})^2\varepsilon/2 - \chi^{(3)}\varepsilon^2/4\varepsilon_0$.

derivation

The displacement field is

$$D = \varepsilon_0 E + P = \varepsilon_0[\frac{\varepsilon}{\varepsilon_0}E + \chi^{(2)}E^2 + \chi^{(3)}E^3 + ...] \tag{10.14}$$

where $\varepsilon/\varepsilon_0 = 1 + \chi^{(1)}$. As we shall see in a moment, the D field is fundamental, so we express E in terms of D by inverting the relation above

$$E = \frac{1}{\varepsilon}D - \frac{\varepsilon_0\chi^{(2)}}{\varepsilon^3}D^2 + \left[\frac{2(\chi^{(2)})^2}{\varepsilon^3} - \frac{\chi^{(3)}}{\varepsilon^2\varepsilon_0}\right]D^3 + ... \tag{10.15}$$

From Jackson, Classical Electrodynamics, section 4.8, we learn that in polarizable materials, the energy in the electric field is

$$W_E = \int d^3r \int_0^D \mathbf{E}\cdot d\mathbf{D}. \tag{10.16}$$

We put in the expansion of E to find

$$W_E = \int d^3r \frac{1}{2\varepsilon} D^2 - \frac{\varepsilon_0 \chi^{(2)}}{3\varepsilon^3} D^3 + \left[\frac{(\chi^{(2)})^2}{2\varepsilon^3} - \frac{\chi^{(3)}}{4\varepsilon^2 \varepsilon_0} \right] D^4 + ... \tag{10.17}$$

When we add the contribution from the magnetic field (we assume the material is nonmagnetic so that $\mu = \mu_0$), we get a Hamiltonian for the field including the effect of the material

$$H_{EM} = W_E + \frac{1}{2\mu_0} \int d^3r B^2. \tag{10.18}$$

This can be written as the linear-optics hamiltonian plus a perturbation $\hat{H}_{EM} = \hat{H}_0 + \hat{H}'$,

$$\hat{H}_0 = \frac{1}{2\varepsilon} \int d^3r \hat{D}^2 + \frac{1}{2\mu_0} \int d^3r \hat{B}^2 = \sum_k \hbar \omega_k (\hat{a}_k^\dagger \hat{a}_k + 1/2) \tag{10.19}$$

$$\hat{H}' = \int d^3r [- \frac{\varepsilon_0 \chi^{(2)}}{3\varepsilon^3} D^3 + \left[\frac{(\chi^{(2)})^2}{2\varepsilon^3} - \frac{\chi^{(3)}}{4\varepsilon^2 \varepsilon_0} \right] D^4 + ...]. \tag{10.20}$$

Note that the frequencies $\omega_k = ck/n = ck/\varepsilon^{1/2}$ already include the first-order susceptibility $\chi^{(1)}$.

10.2.1 aside

Note that if we had used the expansion of D, we would have arrived at a different result, namely

$$\hat{H}_0 = \frac{\varepsilon_0}{2} \int d^3r \hat{E}^2 + \frac{1}{2\mu_0} \int d^3r \hat{B}^2 \quad \text{(wrong!)} \tag{10.21}$$

$$\hat{H}' = \varepsilon_0 \int d^3r [\frac{1}{2} \chi^{(1)} \hat{E}^2 + \frac{2}{3} \chi^{(2)} \hat{E}^3 + \frac{3}{4} \chi^{(3)} \hat{E}^4 + ...] \quad \text{(wrong!)} \tag{10.22}$$

This appears in some early papers, and is clearly not correct. For example, the energy, and thus frequency ω increases with increasing $\chi^{(1)}$. This is contrary to the known behaviour $\omega = ck/(1 + \chi^{(1)})^{1/2}$. But why do we use the expression based on D instead? We go back to the dynamics, now described by the macroscopic Maxwell equations

$$\begin{aligned} \nabla \cdot \mathbf{D} &= 0 && (10.23) \\ \nabla \cdot \mathbf{B} &= 0 && (10.24) \\ \nabla \times \mathbf{E} &= -\frac{\partial \mathbf{B}}{\partial t} && (10.25) \\ \nabla \times \mathbf{H} &= \frac{\partial \mathbf{D}}{\partial t} && (10.26) \end{aligned}$$

These are quantized the same way as the vacuum equations, namely we introduce the vector potential $\mathbf{A}$ through

$$\begin{aligned} \mathbf{B} &= \nabla \times \mathbf{A} \\ \mathbf{D} &= -\frac{\partial \mathbf{A}}{\partial t}. \end{aligned} \tag{10.27}$$

Here $\mathbf{D} = \varepsilon_0(1+\chi^{(1)})\mathbf{E}$ is the linear part of the displacement field. As before $\mathbf{A}$ obeys a wave equation. This means that when we quantize $\mathbf{A}$, its canonical conjugate is $-\mathbf{D}$, not $-\mathbf{E}$. Following the same quantization procedure, we find

$$\hat{\mathbf{D}}(\mathbf{r},t) = i\sum_{k,\alpha}\sqrt{\frac{\hbar\varepsilon\omega_k}{2L^3}}\left(\mathbf{e}_\alpha \hat{a}_{k,\alpha}e^{i\mathbf{k}\cdot\mathbf{r}}e^{-i\omega_k t} - \mathbf{e}_\alpha^* \hat{a}^\dagger_{k,\alpha}e^{-i\mathbf{k}\cdot\mathbf{r}}e^{i\omega_k t}\right). \tag{10.28}$$

Note that the effect of $\chi^{(1)}$ is present in this expression twice: in $\varepsilon = \varepsilon_0(1+\chi^{(1)})$ and also in $\omega_k = ck/n = ck/\sqrt{1+\chi^{(1)}}$. Now the $\hat{\mathbf{E}}$ field is

$$\hat{\mathbf{E}}(\mathbf{r},t) = i\sum_{k,\alpha}\sqrt{\frac{\hbar\omega_k}{2\varepsilon L^3}}\left(\mathbf{e}_\alpha \hat{a}_{k,\alpha}e^{i\mathbf{k}\cdot\mathbf{r}}e^{-i\omega_k t} - \mathbf{e}_\alpha^* \hat{a}^\dagger_{k,\alpha}e^{-i\mathbf{k}\cdot\mathbf{r}}e^{i\omega_k t}\right). \tag{10.29}$$

As a result, the hamiltonian is

$$\hat{H}^{(1)}_{EM} = \frac{1}{2}\int d^3r\left(\frac{1}{\varepsilon}\hat{D}^2 + \frac{1}{\mu_0}\hat{B}^2\right) = \sum_{k,\alpha}\hbar\omega_k(\hat{a}^\dagger_{k,\alpha}\hat{a}_{k,\alpha} + \frac{1}{2}). \tag{10.30}$$

10.2.2 Phenomenological Hamiltonian

We turn back to the phenomenological hamiltonian, which we write now in terms of the $\hat{E}$ field, because this is conventional.

$$\hat{H}_0 = \frac{\varepsilon}{2}\int d^3r\hat{E}^2 + \frac{1}{2\mu_0}\int d^3r\hat{B}^2 = \sum_k \hbar\omega_k(\hat{a}^\dagger_k\hat{a}_k + 1/2) \tag{10.31}$$

$$\hat{H}' = \int d^3r[-\frac{\varepsilon_0\chi^{(2)}}{3} : \hat{E}^3 : + \left[\frac{(\chi^{(2)})^2\varepsilon}{2} - \frac{\chi^{(3)}\varepsilon^2}{4\varepsilon_0}\right] : \hat{E}^4 : + ...]. \tag{10.32}$$

We have assumed that terms like $\hat{E}^4$ are normal ordered (::). That is, all annihilation operators are to the right of the creation operators. This is justified two ways. First, in the spirit of the Glauber photo-detection theory, we assume that the material begins in or near its ground state, so that the first step of any energy-conserving process is the absorption of a photon. Second, even if anti-normal-ordered terms were present, we would be able to absorb them into lower-order terms in the nonlinear expansion. To illustrate, $\hat{E}^4$ contains $\hat{a}\hat{a}\hat{a}^\dagger\hat{a}^\dagger$, but that can be reduced using commutations

$$\hat{a}\hat{a}\hat{a}^\dagger\hat{a}^\dagger = \hat{a}^\dagger\hat{a}^\dagger\hat{a}\hat{a} + 4\hat{a}^\dagger\hat{a} + 2. \tag{10.33}$$

The $4\hat{a}^\dagger\hat{a}$ term acts the same as the $\chi^{(1)}$ term proportional to $\hat{E}^2$. Since we are using the experimental value for $\chi^{(1)}$, this term has already been accounted for.

10.3 Wave-equations approach

Rather than working with the Hamiltonian, from which the dynamics of the field can be derived, we can work directly with the dynamics themselves, meaning we can try to solve (or to approximate) the Maxwell equations when the nonlinear terms are included. This is described in detail in Appendix **??**. For example, the wave equation for transverse fields is

$$\nabla^2 E_i - \frac{n_i^2}{c^2}\frac{\partial^2}{\partial t^2}E_i = \frac{1}{c^2\varepsilon_0}\frac{\partial^2}{\partial t^2}P_i^{(NL)} \tag{10.34}$$

where $n_i \equiv c\sqrt{\mu_0\varepsilon_{ii}}$ is the refractive index for the field polarized along the i direction. In cases where the diffraction is ignorable, we have the 1D wave equation (Eq. **??**)

$$e^{ik_z z - i\omega t}2ik(\partial_z + \frac{n}{c}\partial_t)\mathcal{E} = \frac{1}{c^2\varepsilon_0}\frac{\partial^2}{\partial t^2}P^{(NL)} \tag{10.35}$$

where $E^{(+)}(\mathbf{x},t) \equiv \mathcal{E}(\mathbf{x},t)\exp[ik_z z - i\omega t]$ defines the envelope operator $\mathcal{E}$ in terms of the carrier wave $\exp[ik_z z - i\omega t]$. If the envelopes are slowly varying in time, we can also drop the ∂_t term, and we note that only components of $P^{(NL)}$ with frequencies close to $\exp[-i\omega t]$ will be important. We thus have

$$\partial_z\mathcal{E} = \frac{-ik}{2n^2\varepsilon_0}e^{-ik_z z + i\omega t}P_\omega^{(NL)} \tag{10.36}$$

where $P_\omega^{(NL)}$ is the component of $P^{(NL)}$ oscillating as $\exp[-i\omega t]$.

We now discuss the special case of parametric down-conversion (or sum-frequency generation. They are in fact the same process with different input fields). The three narrow-band fields, called pump, signal and idler, are at centred at frequencies $\omega_p, \omega_s, \omega_i$ with $\omega_p = \omega_s + \omega_i$. Thus, conversion of signal and idler photons into pump photons (or the reverse) satisfies energy conservation, at least approximately. Furthermore, we assume that inside the material, the wave-vectors satisfy $\mathbf{k}_p = \mathbf{k}_s + \mathbf{k}_i$. This is similar to momentum conservation, and is called the *phase-matching* condition. The total field is

$$E(\mathbf{x},t) = E_p(\mathbf{x},t) + E_s(\mathbf{x},t) + E_i(\mathbf{x},t) \tag{10.37}$$

The individual components obey the wave equations

$$\begin{aligned}
\partial_z\mathcal{E}_p &= \frac{-ik_p}{2n_p^2}e^{-ik_p z + i\omega_p t}P_{\omega_p}^{(NL)} \\
\partial_z\mathcal{E}_s &= \frac{-ik_s}{2n_s^2}e^{-ik_s z + i\omega_s t}P_{\omega_s}^{(NL)} \\
\partial_z\mathcal{E}_i &= \frac{-ik_i}{2n_i^2}e^{-ik_i z + i\omega_i t}P_{\omega_i}^{(NL)}
\end{aligned} \tag{10.38}$$

The nonlinear polarization is

$$P^{(NL)} = \varepsilon_0\chi^{(2)}(E_p + E_s + E_i)^2 \tag{10.39}$$

which contains many terms, proportional to all the pairs of $\{\mathcal{E}_p, \mathcal{E}_s, \mathcal{E}_i, \mathcal{E}_p^\dagger, \mathcal{E}_s^\dagger, \mathcal{E}_i^\dagger\}$. But if we look at the equations above, we see that only terms which vary approximately as $\exp[ik_p z - i\omega_p t]$, $\exp[ik_s z - i\omega_s t]$ or $\exp[ik_i z - i\omega_i t]$ will have a significant effect. Other terms will be rapidly oscillating. We keep three terms, one for each equation

$$\begin{aligned} &\varepsilon_0 \chi^{(2)} \mathcal{E}_s \mathcal{E}_i e^{i(k_s+k_i)z - i(\omega_s+\omega_i)t} \\ &\varepsilon_0 \chi^{(2)} \mathcal{E}_p \mathcal{E}_i^\dagger e^{i(k_p-k_i)z - i(\omega_p-\omega_i)t} \\ &\varepsilon_0 \chi^{(2)} \mathcal{E}_p \mathcal{E}_s^\dagger e^{i(k_p-k_s)z - i(\omega_p-\omega_s)t}. \end{aligned} \tag{10.40}$$

When they are inserted into the wave equations we have

$$\begin{aligned} \partial_z \mathcal{E}_p &= ig\mathcal{E}_s\mathcal{E}_i \\ \partial_z \mathcal{E}_s &= ig\mathcal{E}_p\mathcal{E}_i^\dagger \\ \partial_z \mathcal{E}_i &= ig\mathcal{E}_p\mathcal{E}_s^\dagger \end{aligned} \tag{10.41}$$

where $g \propto \chi^{(2)}$. We assume that the pump is in a strong coherent state $\mathcal{E}_p = \alpha_p$, and does not appreciably change over the length of the crystal. The equations have the form

$$\partial_z \mathcal{E}_p = ig\mathcal{E}_s\mathcal{E}_i \approx 0 \tag{10.42}$$

$$\partial_z \mathcal{E}_s = ig\mathcal{E}_p\mathcal{E}_i^\dagger = ig'\mathcal{E}_i^\dagger \tag{10.43}$$

$$\partial_z \mathcal{E}_i = ig\mathcal{E}_p\mathcal{E}_s^\dagger = ig'\mathcal{E}_s^\dagger \tag{10.44}$$

where $g' = g\alpha_p$. Note that g' depends on the phase of the pump, so we should expect this to be a *phase-sensitive* process. In fact it is, if we re-write these as

$$\partial_z \mathcal{E}_s = |g| e^{i\phi_p} \mathcal{E}_i^\dagger \tag{10.45}$$

$$\partial_z e^{i\phi_p} \mathcal{E}_i^\dagger = |g| \mathcal{E}_s \tag{10.46}$$

we have the solution

$$\mathcal{E}_s(z) = \mathcal{E}_s(0)\cosh(|g|z) + e^{i\phi_p}\mathcal{E}_i^\dagger(0)\sinh(|g|z) \tag{10.47}$$

$$e^{i\phi_p}\mathcal{E}_i^\dagger(z) = e^{i\phi_p}\mathcal{E}_i^\dagger(0)\cosh(|g|z) + \mathcal{E}_s(z)\sinh(|g|z) \tag{10.48}$$

or

$$\mathcal{E}_s(z) = \mathcal{E}_s(0)\cosh(|g|z) + e^{i\phi_p}\mathcal{E}_i^\dagger(0)\sinh(|g|z) \tag{10.49}$$

$$\mathcal{E}_i(z) = \mathcal{E}_i(0)\cosh(|g|z) + e^{i\phi_p}\mathcal{E}_s^\dagger(0)\sinh(|g|z) \tag{10.50}$$

Because we assumed plane wave fields, these expressions apply just to pairs of signal and idler modes, those which satisfy the phase matching condition $\mathbf{k}_s + \mathbf{k}_i = \mathbf{k}_p$. For these modes, we can write the output operators in terms of the input operators as

$$a_{k_s,\text{out}} = a_{k_s,\text{in}}\cosh(|g|L) + a_{k_i,\text{in}}^\dagger \sinh(|g|L) \tag{10.51}$$

$$a_{k_i,\text{out}} = a_{k_i,\text{in}}\cosh(|g|L) + a_{k_s,\text{in}}^\dagger \sinh(|g|L). \tag{10.52}$$

This agrees precisely with the effect of the two-mode squeeze operator

$$S_2(G) = \exp[G^* \hat{a}_+ \hat{a}_- - G\hat{a}_+^\dagger \hat{a}_-^\dagger] \tag{10.53}$$

namely

$$S_2^\dagger(G)\hat{a}_\pm S_2(G) = \hat{a}_\pm \cosh r - \hat{a}_\mp^\dagger e^{i\theta} \sinh r \tag{10.54}$$

if we take $G \equiv r\exp[i\theta] = |g|\exp[i\phi_p]$. If the two modes are the same, this becomes the ordinary squeeze operator

$$S(\varepsilon) \equiv \exp[\frac{1}{2}\varepsilon^* \hat{a}^2 - \frac{1}{2}\varepsilon(\hat{a}^\dagger)^2] \tag{10.55}$$

with $\varepsilon = 2|g|\exp[i\phi_p]$.

10.4 Parametric down-conversion

If we have a crystal with nonzero $\chi^{(2)}$, we can pump it with a blue laser and produce pairs of red photons. We assume that the field starts in the state $|\phi_0\rangle = |\alpha\rangle_p |0\rangle_s |0\rangle_i$ where p, s, i indicate the pump, signal and idler, respectively. The only part of the hamiltonian that is interesting to us is $H' = -\varepsilon_0/3 \int d^3r \chi^{(2)} \hat{E}_s^{(-)} \hat{E}_i^{(-)} \hat{E}_p^{(+)}$, which consumes a pump photon and produces a signal and idler photon. We do the calculation in the interaction picture, although it can equally well be done in the Heisenberg picture[1]. The state evolves as

$$|\phi\rangle = e^{\frac{-iH't}{\hbar}} |\phi_0\rangle = |\phi_0\rangle + \frac{i\varepsilon_0}{\hbar} \int_0^t dt' \int d^3r \chi^{(2)} \hat{E}_s^{(-)} \hat{E}_i^{(-)} \hat{E}_p^{(+)} |\phi_0\rangle + O(2) \tag{10.56}$$

where $O(2)$ indicates higher-order terms. We keep just the first-order term, and expand in modes

$$|\phi\rangle \propto \int_0^t dt' \sum_{k_p k_s k_i} \int d^3r \chi^{(2)}(\mathbf{r}) \hat{a}_{k_s}^\dagger \hat{a}_{k_i}^\dagger \hat{a}_{k_p} e^{i(\mathbf{k}_p - \mathbf{k}_s - \mathbf{k}_i)\cdot\mathbf{r}} e^{-i(\omega_p - \omega_s - \omega_i)t'} |\phi_0\rangle \,. \tag{10.57}$$

If we assume that $\chi^{(2)}(\mathbf{r})$ is from a rectangular crystal of dimensions L_x, L_y, L_z, we can do the spatial and temporal integrals to get

$$\begin{aligned} |\phi\rangle \quad \propto \quad & \frac{t\chi^{(2)} L_x L_y L_z}{16} \sum_{k_p k_s k_i} \mathrm{sinc}[\Delta\omega t/2] \\ & \times \mathrm{sinc}[\Delta\mathbf{k}_x L_x/2]\mathrm{sinc}[\Delta\mathbf{k}_y L_x/2]\mathrm{sinc}[\Delta\mathbf{k}_z L_z/2] \hat{a}_{k_s}^\dagger \hat{a}_{k_i}^\dagger \hat{a}_{k_p} |\phi_0\rangle \,. \end{aligned} \tag{10.58}$$

where $\Delta\omega = (\omega_p - \omega_s - \omega_i)$ and $\Delta\mathbf{k} = (\mathbf{k}_p - \mathbf{k}_s - \mathbf{k}_i)$ and $\mathrm{sinc}(x) = \sin(x)/x$ is the sinc function. Note that the sinc function expresses the well-known condition of phase-matching, and contains the wave-vectors for pump, signal and idler *in the material*. We assume that the pump is a plane wave propagating in the $+z$ direction, so that only one term contributes to the sum over

[1] Hong, C. K. and Mandel, L. (1985) Theory of parametric frequency down conversion of light. Physical Review A 31, 2409-18.

k_p. Then $\hat{a}_{k_p}|\phi\rangle = \alpha_p|\phi\rangle$. We assume that t is large so we can replace the $t\mathrm{sinc}(\Delta\omega t)$ with a delta function $\delta(\Delta\omega)$. Similarly, we assume that the crystal is relatively wide ($L_x, L_y \gg L_z$) so that $L_{x,y}\mathrm{sinc}(\Delta\mathbf{k}_{x,y}L_{x,y}) \to \delta(\Delta\mathbf{k}_{x,y})$. We then find

$$|\phi\rangle \propto \sum_{k_s k_i} \delta(\Delta\omega)\delta(\Delta\mathbf{k}_x)\delta(\Delta\mathbf{k}_y)\mathrm{sinc}[\Delta\mathbf{k}_z L_z/2]\hat{a}^\dagger_{k_s}\hat{a}^\dagger_{k_i}|\phi_0\rangle\,. \tag{10.59}$$

This state has interesting properties. For one thing, the signal and idler are correlated, due to the phase matching and energy conservation. The sum of their momenta is $\mathbf{k}_s + \mathbf{k}_i \approx \mathbf{k}_p$, but their individual momenta are highly uncertain. In fact, they are entangled, because the expression $\mathrm{sinc}[\Delta\mathbf{k}_z L_z/2]\hat{a}^\dagger_{k_s}\hat{a}^\dagger_{k_i}$ does not factor. The signal and idler are also correlated in time. We compute

$$\begin{aligned} G^{(2)}_{s,i}(t, t+\tau) &\equiv \left\langle \hat{E}^{(-)}_s(t)\hat{E}^{(-)}_i(t+\tau)\hat{E}^{(+)}_i(t+\tau)\hat{E}^{(+)}_s(t)\right\rangle \\ &= |\langle\phi_0|\,\hat{E}^{(+)}_i(t+\tau)\hat{E}^{(+)}_s(t)\,|\phi\rangle\,|^2 \equiv |\Psi_{s,i}(t, t+\tau)|^2. \end{aligned} \tag{10.60}$$

The function $\Psi_{s,i}(t, t+\tau)$ is sometimes called the "two-photon wave-function" because it is an amplitude that when squared gives the probability of finding two photons. Later it will be useful to have this amplitude, so we compute it, rather than working with $G^{(2)}$.

$$\begin{aligned} \Psi_{s,i}(t, t+\tau) &\propto \sum_{k_s k_i} \langle\phi_0|\,\hat{a}_{k_i}\hat{a}_{k_s} e^{-i(\omega_s+\omega_i)t} e^{i(\mathbf{k}_s\cdot\mathbf{r}_s + \mathbf{k}_i\cdot\mathbf{r}_i)} e^{-i\omega_i\tau} \\ &\quad \times \sum_{k'_s k'_i} \delta(\Delta\omega')\delta(\Delta\mathbf{k}_x)\delta(\Delta\mathbf{k}_y)\mathrm{sinc}[\Delta\mathbf{k}'_z L_z/2]\hat{a}^\dagger_{k'_s}\hat{a}^\dagger_{k'_i}|\phi_0\rangle \\ &= \sum_{k_s k_i} e^{-i(\omega_s+\omega_i)t} e^{i(\mathbf{k}_s\cdot\mathbf{r}_s + \mathbf{k}_i\cdot\mathbf{r}_i)} e^{-i\omega_i\tau} \\ &\quad \times\delta(\Delta\omega)\delta(\Delta\mathbf{k}_x)\delta(\Delta\mathbf{k}_y)\mathrm{sinc}[\Delta\mathbf{k}_z L_z/2]. \end{aligned} \tag{10.61}$$

The delta functions indicate that the properties of the emitted photons are constrained. In fact, the three delta functions are sufficient to completely determine $\mathbf{k}_s$ given $\mathbf{k}_i$ (or vice versa). Furthermore, the sinc function means that some values of $\mathbf{k}_i$ contribute more than others. If we parametrize the direction of emission as (θ_i, ϕ_i) and (θ_s, ϕ_s). Then we are free to choose ϕ_i and *either* ω_i or θ_i. Then the remaining degrees of freedom are determined. We assume the detectors are placed to collect $\phi_i = 0$, and that filters transmit ω_i with an efficiency $f(\omega_i)^2$.

We assume that there are some wave-vectors $(\mathbf{k}_{s,0}, \mathbf{k}_{i,0})$ which perfectly satisfy both phase-matching $\mathbf{k}_{s,0} + \mathbf{k}_{i,0} = \mathbf{k}_{p,0}$ and energy conservation $\omega_{s,0} + \omega_{i,0} \equiv \omega_{k_{s,0}} + \omega_{k_{i,0}} = \omega_p$. We expand around these centre values as

$$\mathbf{k}_i = \mathbf{k}_{i,0} + \left(\frac{\partial\mathbf{k}_i}{\partial\omega_i}\right)_{\Delta\mathbf{k}_\perp=0}\delta\omega_i \tag{10.62}$$

$$\mathbf{k}_s = \mathbf{k}_{s,0} + \left(\frac{\partial\mathbf{k}_s}{\partial\omega_s}\right)_{\Delta\mathbf{k}_\perp=0}\delta\omega_s = \mathbf{k}_{s,0} - \left(\frac{\partial\mathbf{k}_s}{\partial\omega_s}\right)_{\Delta\mathbf{k}_\perp=0}\delta\omega_i \tag{10.63}$$

[2]Note that, since ω_i and θ_i are not independent, $f(\omega_i)$ could reflect either frequency or angular filtering, or both.

where $\omega_s = \omega_{s,0} + \delta\omega_s$ and $\omega_i = \omega_{i,0} + \delta\omega_i$ and by energy conservation $\delta\omega_s = -\delta\omega_i$.

We can then write the two-photon wave-function as

$$\begin{aligned}\Psi_{s,i}(t, t+\tau) \quad &\propto \quad e^{-i\omega_p t} e^{i(\mathbf{k}_{s,0}\cdot\mathbf{r}_s + \mathbf{k}_{i,0}\cdot\mathbf{r}_i)} \sum_{k_i} e^{i(-\partial\mathbf{k}_{s,z}/\partial\omega_s \cdot \mathbf{r}_s \delta\omega_i + \partial\mathbf{k}_{s,z}/\partial\omega_i \cdot \mathbf{r}_i \delta\omega_i)} \\ &\quad \times e^{-i\omega_i\tau} f(\omega_i) \mathrm{sinc}[\delta\omega_i (v_{g,s,z}^{-1} - v_{g,i,z}^{-1}) L_z/2] \, |\phi_0\rangle \end{aligned} \tag{10.64}$$

where

$$\begin{aligned} v_{g,s,z}^{-1} \quad &\equiv \quad \left(\frac{\partial \mathbf{k}_{s,z}}{\partial \omega_s}\right)_{\Delta\mathbf{k}_\perp = 0} \\ v_{g,i,z}^{-1} \quad &\equiv \quad \left(\frac{\partial \mathbf{k}_{i,z}}{\partial \omega_i}\right)_{\Delta\mathbf{k}_\perp = 0} \end{aligned} \tag{10.65}$$

are the inverse group velocities in the forward direction. We can drop the global phase $\exp[i(\mathbf{k}_{s,0}\cdot\mathbf{r}_s + \mathbf{k}_{i,0}\cdot\mathbf{r}_i)]$ without changing $G^{(2)}$. The complicated-looking phase factor $\exp[i(-\partial\mathbf{k}_{s,z}/\partial\omega_s \cdot \mathbf{r}_s\delta\omega_i + \partial\mathbf{k}_{s,z}/\partial\omega_i \cdot \mathbf{r}_i\delta\omega_i)]$ is identically 1 if the detectors are placed the same time-of-flight away from the crystal (i.e., photons created simultaneously at the exit face of the crystal arrive simultaneously to the two detectors). We assume this is the case, so we have

$$\Psi_{s,i}(t, t+\tau) \propto e^{-i\omega_p t} \sum_{k_i} e^{-i\omega_i\tau} f(\omega_i) \mathrm{sinc}[\delta\omega_i (v_{g,s,z}^{-1} - v_{g,i,z}^{-1}) L_z/2] \, |\phi_0\rangle \, . \tag{10.66}$$

We can replace the sum over k_i with an integral $\sum_{k_i} e^{-i\omega_i\tau} \to e^{-i\omega_{k_{0,i}}\tau} \int_{-\infty}^{\infty} d\delta\omega_i e^{-i\delta\omega_i}$ to get

$$\Psi_{s,i}(t, t+\tau) \propto e^{-i\omega_p t} e^{-i\omega_{k_{0,i}}\tau} \int_{-\infty}^{\infty} d\delta\omega_i \, e^{-i\delta\omega_i\tau} f(\delta\omega_i) \mathrm{sinc}[\delta\omega_i (v_{g,s,z}^{-1} - v_{g,i,z}^{-1}) L_z/2]. \tag{10.67}$$

This is the Fourier transform of the sinc function times the filter function, in general a convolution of the filter's time-response $F(t)$ and the Fourier transform of the sinc function, the rectangular function,

$$\Psi_{s,i}(t, t+\tau) \propto F(\tau) \otimes \begin{cases} 0 & \tau < 0 \\ \exp[-i\omega_{s,0} t] \exp[-i\omega_{i,0}(t+\tau)]/\delta_{tt} & 0 < \tau < \delta_{tt} \\ 0 & \tau > \delta_{tt} \end{cases} \tag{10.68}$$

where $\delta_{tt} \equiv (v_{g,s,z}^{-1} - v_{g,i,z}^{-1}) L_z$ is the difference in transit times through the crystal for the signal and idler photons. In many cases, one or the other contribution will dominate. For example, in degenerate type-I phase-matching, the group velocities of signal and idler are the same, and thus $\delta_{tt} = 0$. Then the contribution of the filter is then all-important. For collinear type-II phase-matching, it is usually the crystal thickness that dominates.

This crystal's contribution has a very simple explanation. At any point in the crystal, a pump photon can down-convert to become a signal-idler pair. When these are created, they are produced at the same time and place. Because they may have different group velocities, their arrival time at the detectors can differ by up to δ_{tt}, the transit time difference for passing through the whole crystal.

Chapter 11

Quantum optics with atomic ensembles

11.1 Atoms

The physics *inside* of an atom is very rich: an atom is a strongly-interacting system of relativistic electrons trapped in the potential of the nucleus, which contributes interesting features of its own. For our purposes, however, we are interested only in the interaction of the atom with light fields, and from the *outside*, an atom appears very simple. We assume a collection of identical atoms, i.e., that each atom has the same internal characteristics. The state of each atom can be expanded in energy eigenstates $|\phi_1\rangle, |\phi_2\rangle, ...$ with energies $\hbar\omega_1, \hbar\omega_2, ...$ The centre-of-mass motion of the atom is that of a free particle[1], so that the atomic Hamiltonian for the ith atom is

$$H_{at,i} = \hbar \sum_j \omega_j |\phi_j\rangle_i \langle\phi_j|_i + \frac{p_{at,i}^2}{2m} \tag{11.1}$$

where $p_{at,i}$ is the atomic momentum operator and m is the atomic mass. External fields can cause transitions among these states, usually by electric dipole transitions (other transitions are much weaker). This is described by an interaction Hamiltonian

$$H_{int,i} = -\mathbf{E}(\mathbf{x}_{at,i}) \cdot \mathbf{d}_i \tag{11.2}$$

where $\mathbf{E}$ is the electric field, $\mathbf{x}_{at,i}$ is the atomic position operator, and $\mathbf{d}_i$ is the electric dipole operator. The dipole operator can be written in the form

$$\mathbf{d}_i = |\phi_j\rangle_i \mathbf{d}_{jk} \langle\phi_k|_i \tag{11.3}$$

where $\mathbf{d}_{jk}$ is called the transition dipole moment or dipole matrix element, between states k and j.

[1] We could easily add a potential to this Hamiltonian to describe external forces, e.g. gravity. Note that optically created potentials, such as optical dipole potentials, would be produced by the interaction Hamiltonian that follows.

Each atom contributes to the total Hamiltonian, so that

$$H = H_{EM} + H_{AT} + H_{INT} \tag{11.4}$$

where H_{EM} is the Hamiltonian for the field and

$$\begin{aligned} H_{AT} &= \sum_i H_{at,i} \\ H_{INT} &= \sum_i H_{int,i}. \end{aligned} \tag{11.5}$$

11.1.1 Rotating-wave approximation

The dipole interaction Hamiltonian H_{int} contains $\mathbf{E} = \mathbf{E}^{(+)} + \mathbf{E}^{(-)}$, which can either raise or lower the number of photons in the field. Similarly, $\mathbf{d}$ can either raise or lower the energy of the atom. For convenience, we write $\mathbf{d} = \mathbf{d}^{(+)} + \mathbf{d}^{(-)}$ where

$$\mathbf{d}^{(+)} \equiv \sum_{\omega_j < \omega_k} |\phi_j\rangle \, \mathbf{d}_{jk} \, \langle \phi_k| \tag{11.6}$$

and $\mathbf{d}^{(-)} \equiv [\mathbf{d}^{(+)}]^\dagger$. Note that $\mathbf{d}^{(+)}$ lowers the energy of the atom (as $\mathbf{E}^{(+)}$ lowers the energy of the field). The "rotating wave approximation" (the name comes from nuclear magnetic resonance) is made by dropping the "counter-rotating terms," i.e., those which would either raise both field and atom energy, or lower both[2]. Dropping these terms, we have

$$\begin{aligned} H_{int,i} &= -\mathbf{E}(\mathbf{x}_{at,i}) \cdot \mathbf{d}_i \\ &\approx -\left[\mathbf{E}^{(+)}(\mathbf{x}_{at,i}) \cdot \mathbf{d}_i^{(-)} + \mathbf{E}^{(-)}(\mathbf{x}_{at,i}) \cdot \mathbf{d}_i^{(+)} \right]. \end{aligned} \tag{11.7}$$

11.1.2 First-order light-atom interactions

In lowest order, a photon is absorbed while the atom makes a transition to a higher-energy state, or the reverse: a photon is emitted and the atom drops to a lower-energy state. These processes will only occur when the photon has the same energy as the transition, to within uncertainties due to the finite lifetime of the atomic states and the photon coherence time. These processes are fundamental to photo-detection and to laser amplification. The excited atom spontaneously emits in a random direction, which in any practical situation implies loss of information about the state of the field, or equivalently introduction of noise. For this reason, most proposals for the manipulation of quantum light use higher-order processes, and avoid exciting the atomic upper levels[3].

[2] Because these terms do not conserve energy, they do not contribute to first-order processes. They can contribute to higher-order processes. A famous example is the Lamb shift, which corresponds to $\mathbf{E}^{(-)}\mathbf{d}^{(-)}$ (emission of a photon and transition to a higher level) followed by $\mathbf{E}^{(+)}\mathbf{d}^{(+)}$. Even so, in most situations there is some other process, allowed by the RWA, which is dominant.

[3] An important exception is the control of absorption and spontaneous emission that is possible in cavity QED. When an atom is placed within a high-finesse cavity, its interaction with the modes of the cavity is much stronger than with other modes, and the emission into random directions can be greatly reduced. This strategy has been successfully pursued with neutral atoms and ions, but remains technically very challenging.

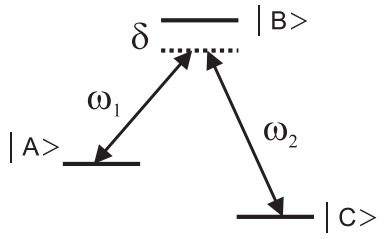

Figure 11.1: Atomic levels in a "lambda" atom.

11.1.3 Second-order light-atom interactions

Light which is not resonant with a transition cannot be absorbed, but it can participate in higher-order processes such Raman scattering, in which a photon is absorbed on one transition and simultaneously emitted on a different transition. This process appears in second-order perturbation theory, as described in Appendix B. If the components of the light have frequency ω, and the lower and upper atomic states have energies $\hbar\omega_{l,k}$, respectively, then the detuning is $\delta_{kl} \equiv \omega - (\omega_k - \omega_l)$. When the bandwidth of the light is small compared to the detuning, the first-order terms can be ignored, and the effective Hamiltonian

$$\begin{aligned} H_{\text{eff},i} &= \sum_{\substack{k \in \{\text{upper}\} \\ j,l \in \{\text{lower}\}}} \mathbf{E}^{(-)}(\mathbf{x}_i) \cdot |\phi_j\rangle \frac{\mathbf{d}_{jk}\mathbf{d}_{kl}}{\delta_{kl}} \langle\phi_l| \cdot \mathbf{E}^{(+)}(\mathbf{x}_i) + H.c. \\ &\equiv -\mathbf{E}^{(-)}(\mathbf{x}_i) \cdot \overleftrightarrow{\alpha}_i \cdot \mathbf{E}^{(+)}(\mathbf{x}_i) + H.c. \end{aligned} \tag{11.8}$$

describes the light-atom interaction, replacing $H_{int,i}$ in the interaction Hamiltonian. Here $\overleftrightarrow{\alpha}$ is a tensor operator which describes the polarizability

$$\overleftrightarrow{\alpha} \equiv \sum_{\substack{k \in \{\text{upper}\} \\ j,l \in \{\text{lower}\}}} |\phi_j\rangle \frac{\mathbf{d}_{jk}\mathbf{d}_{kl}}{\delta_{kl}} \langle\phi_l| . \tag{11.9}$$

11.2 Atomic ensembles

The interaction between a single atom and a single photon is typically very weak. Collections of many identical atoms, "atomic ensembles," naturally have a much stronger effect. What is perhaps surprising is that atomic ensembles can behave like simple quantum systems, much like single atoms or single modes of a light field. The general strategy is to find a degree of freedom of the entire ensemble which interacts with the light field, and to study the behaviour of that collective degree of freedom. Two examples are "collective continuous variables," for example the total spin operator of the ensemble, and "collective excitations," which are something like spin-wave quasi-particles, and can be described with creation and annihilation operators as if they were photons.

11.2.1 collective excitations

To show the collective excitations approach as simply as possible, we consider an atom with only two lower levels and one upper level as in figure 10.1. We assume that the lower levels are coupled to the upper levels by different parts of the **E** field, either because the transition have different polarizations, or because they have different energies, or both. We write these parts of the field as E_1, E_2. We note that the interaction Hamiltonian is

$$H_{INT} = -\alpha_{CA} \sum_i E_2^{(-)}(\mathbf{x}_i) \left|C\right\rangle_i \left\langle A\right|_i E_1^{(+)}(\mathbf{x}_i) + H.c. \tag{11.10}$$

where $\alpha_{CA} = d_{CB} d_{BA}/\delta$. The structure of the interaction is clear: a Raman transition is the movement of an atom from state A to C and the scattering of a photon from field E_1 to E_2. The reverse process is included in the Hermitian conjugate term.

We take a particular atomic state, $|A, A, \ldots, A\rangle = |A\rangle^{\otimes N} \equiv |0\rangle_{\text{Atoms}}$ as a reference state. Other states can be made from this by application of the transition operators $T_i^\dagger \equiv \left|C\right\rangle_i \left\langle A\right|_i$. $T^\dagger$ is something like a creation operator, it creates one atom that is not in the initial state. Note that this is not a bosonic operator: $[T_i, T_j^\dagger] = \left|A\right\rangle_i \left\langle C\right|_i \left|C\right\rangle_j \left\langle A\right|_j = \left|A\right\rangle_i \left\langle A\right|_j \delta_{ij} \neq 1$.

$$H_{INT} = -\alpha_{CA} \sum_i E_2^{(-)}(\mathbf{x}_i) T_i^\dagger E_1^{(+)}(\mathbf{x}_i) + H.c. \tag{11.11}$$

Note that this effective Hamiltonian has a similar structure to that of parametric down-conversion. One field loses a photon (by $E_1^{(+)}(\mathbf{x}_i)$) while two excitations are created (by $E_2^{(-)}(\mathbf{x}_i)$ and $T_i^\dagger$). The whole process is local: the creation and annihilation occur where the atom is.

It is convenient to express this in momentum space, using the expansion

$$E^{(+)}(\mathbf{x}, t) = i \sum_{\mathbf{k}} \sqrt{\frac{\hbar\omega_{\mathbf{k}}}{2\varepsilon_0 V}} a_{\mathbf{k}}(t) e^{i\mathbf{k}\cdot\mathbf{x}} \equiv ig \sum_{\mathbf{k}} a_{\mathbf{k}}(t) e^{i\mathbf{k}\cdot\mathbf{x}} \tag{11.12}$$

$$H_{INT} = -\alpha_{CA} g_1 g_2^* \sum_i \sum_{\mathbf{k}_1, \mathbf{k}_2} a_{\mathbf{k}_2}^\dagger a_{\mathbf{k}_1} T_i^\dagger e^{i(\mathbf{k}_1 - \mathbf{k}_2)\cdot\mathbf{x}_i} + H.c. \tag{11.13}$$

This is very suggestive, and we define a creation operator for a collective excitation as

$$A_{\mathbf{k}}^\dagger \equiv N^{-1/2} \sum_i T_i^\dagger e^{i\mathbf{k}\cdot\mathbf{x}_i} \tag{11.14}$$

The interaction Hamiltonian is now

$$H_{INT} = -\alpha_{CA} g_1 g_2^* N^{-1/2} \sum_{\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_A} a_{\mathbf{k}_2}^\dagger A_{\mathbf{k}_A}^\dagger a_{\mathbf{k}_1} \delta_{\mathbf{k}_1, \mathbf{k}_2 + \mathbf{k}_A} + H.c. \tag{11.15}$$

This has a great similarity to the interaction Hamiltonian for parametric down-conversion, except that one "mode" describes the state of excitation of the atoms. We note that the creation

operator $A^\dagger_{\mathbf{k}}$ does not excite any particular atom. If we allow this to act on the reference state, we have

$$\begin{aligned} A^\dagger_{\mathbf{k}} \left|0\right\rangle_{\text{Atoms}} &= N^{-1/2} \sum_i T^\dagger_i e^{i\mathbf{k}\cdot\mathbf{x}_i} \left|A, A \ldots, A\right\rangle \\ &= N^{-1/2} \Big\{ e^{i\mathbf{k}\cdot\mathbf{x}_1} \left|C, A \ldots, A\right\rangle + e^{i\mathbf{k}\cdot\mathbf{x}_2} \left|A, C \ldots, A\right\rangle \\ &\quad + \ldots + e^{i\mathbf{k}\cdot\mathbf{x}_N} \left|A, A \ldots, C\right\rangle \Big\} \end{aligned} \tag{11.16}$$

We see why this is called a collective excitation: there is a single excitation (one atom in state C), but the amplitude is spread evenly over all the possible atoms. There is a sense in which this state has a momentum $\mathbf{k}$. For example, under translations of the coordinate system by $\delta\mathbf{x}$, the state changes by a global phase $\exp[i\mathbf{k}\cdot\delta\mathbf{x}]$. In other ways it is not exactly like a momentum state of a particle. For example, two states of different momentum are not in general orthogonal

$$\left\langle 0\right| A_{\mathbf{k}} A^\dagger_{\mathbf{k}'} \left|0\right\rangle = N^{-1} \sum_i e^{i(\mathbf{k}'-\mathbf{k})\cdot\mathbf{x}_i}. \tag{11.17}$$

In the case where the $\mathbf{x}_i$ are random (a gas of atoms),

$$\begin{aligned} \left|\left\langle 0\right| A_{\mathbf{k}} A^\dagger_{\mathbf{k}'} \left|0\right\rangle\right|^2 &= N^{-2} \sum_i \left| e^{i(\mathbf{k}'-\mathbf{k})\cdot\mathbf{x}_i} \right|^2 + N^{-2} \sum_{i\neq j} e^{-i(\mathbf{k}'-\mathbf{k})\cdot\mathbf{x}_i} e^{i(\mathbf{k}'-\mathbf{k})\cdot\mathbf{x}_j} \\ &= N^{-1} + N^{-2} X \end{aligned} \tag{11.18}$$

where X is the sum of random complex exponentials. X has zero average and RMS fluctuation of order N. We see that for large numbers of atoms, the excitations become approximately orthogonal.

application: DLCZ photon source

Consider the following scenario: we start with the atoms in state $\left|0\right\rangle_{\text{Atoms}}$ and turn on a classical field E_1 in the k_1 direction. The interaction

$$H_{INT} = -\alpha_{CA} g_1 g_2^* N^{-1/2} \sum_{\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_A} a^\dagger_{\mathbf{k}_2} A^\dagger_{\mathbf{k}_A} a_{\mathbf{k}_1} \delta_{\mathbf{k}_1,\mathbf{k}_2+\mathbf{k}_A} + H.c. \tag{11.19}$$

will now act to produce pairs of excitations $a^\dagger_{\mathbf{k}_2} A^\dagger_{\mathbf{k}_A}$ for any momentum pair that satisfies $\mathbf{k}_2+\mathbf{k}_A = \mathbf{k}_1$. If we detect a photon in the mode $\mathbf{k}_2$, we can then infer the presence of an atomic excitation $A^\dagger_{\mathbf{k}_1-\mathbf{k}_2}$. This detected photon is called a "herald," something that signals the presence of another, un-observed object. The collective excitation is stationary and reasonably stable (its lifetime is not limited by spontaneous emission, but rather by the time it takes the atoms to move and change the $\exp[i(\mathbf{k}_1-\mathbf{k}_2)\cdot x]$ factors). We can turn off the pump light in $\mathbf{k}_1$ and leave the collective excitation in the ensemble. After waiting some variable period of time, we turn on a classical field E_2 in the $\mathbf{k}'_2$ direction, with $\mathbf{k}'_2$ anti-parallel to $\mathbf{k}_1$. This causes a Raman transition which returns the atom to the state $\left|A\right\rangle$. More precisely, the term $a_{\mathbf{k}'_2} A_{\mathbf{k}_1-\mathbf{k}_2} a^\dagger_{\mathbf{k}_1-\mathbf{k}_2+\mathbf{k}'_2}$ produces

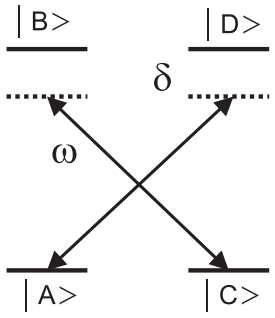

Figure 11.2: Atomic levels in a "spin-1/2" or "X" atom.

an outgoing photon with momentum $\mathbf{k}_{out} = \mathbf{k}_1 - \mathbf{k}_2 + \mathbf{k}_2'$. Note that there is no uncertainty about the second process: the classical field is fixed by the experimenter, and the momentum state of the collective excitation is known from the heralding photon. If the transitions are of approximately the same energy, such that $\mathbf{k}_1 + \mathbf{k}_2' \approx 0$, then $\mathbf{k}_{out} \approx -\mathbf{k}_2$, and the output photon is emitted in roughly the opposite direction from the trigger photon.

11.2.2 collective continuous variables

If we return to the interaction Hamiltonian,

$$H_{\text{eff},i} = -\mathbf{E}^{(-)}(\mathbf{x}_i) \cdot \overleftrightarrow{\alpha}_i \cdot \mathbf{E}^{(+)}(\mathbf{x}_i) + H.c. \tag{11.20}$$

we can re-formulate this in terms of macroscopic quantum variables for the light and atoms. To be concrete, we consider an atom with the structure of Figure 10.2. Again, we consider two fields E_1, E_2, this time distinguished only by their polarizations. Because the allowed transitions couple only A to D and C to B, and each with only one field, we have

$$\begin{aligned} H_{\text{eff},i} &= -\frac{|\mathbf{d}_{AD}|^2}{\delta} |A\rangle_i \langle A|_i E_1^{(-)}(\mathbf{x}_i) E_1^{(+)}(\mathbf{x}_i) - \frac{|\mathbf{d}_{CB}|^2}{\delta} |C\rangle_i \langle C|_i E_2^{(-)}(\mathbf{x}_i) E_2^{(+)}(\mathbf{x}_i) \\ &= -\alpha_0 \left(|A\rangle_i \langle A|_i E_1^{(-)}(\mathbf{x}_i) E_1^{(+)}(\mathbf{x}_i) + |C\rangle_i \langle C|_i E_2^{(-)}(\mathbf{x}_i) E_2^{(+)}(\mathbf{x}_i) \right). \end{aligned} \tag{11.21}$$

From the perspective of the atoms, the light causes AC-stark shifts: the atomic level is shifted in proportion to the intensity of the light driving the transition. From the perspective of the light, the atom changes the refractive index for one polarization or the other, depending on the state of the atom.

We note that the field can be described in terms of the Stokes operators

$$\begin{aligned} S_0 &\equiv E_1^{(-)} E_1^{(+)} + E_2^{(-)} E_2^{(+)} \\ S_x &\equiv E_1^{(-)} E_2^{(+)} + E_2^{(-)} E_1^{(+)} \\ S_y &\equiv i \left(E_1^{(-)} E_2^{(+)} - E_2^{(-)} E_1^{(+)} \right) \\ S_z &\equiv E_1^{(-)} E_1^{(+)} - E_2^{(-)} E_2^{(+)} \end{aligned} \tag{11.22}$$

Also, we can write "spin" operators $\mathbf{j}_i = (j_{i,x}, j_{i,y}, j_{i,z})^T$ with

$$\begin{aligned} j_{i,0} &= \frac{1}{2}\left(|C\rangle_i \langle C|_i + |A\rangle_i \langle A|_i\right) = \frac{1}{2} \\ j_{i,x} &= \frac{1}{2}\left(|A\rangle_i \langle C|_i + |C\rangle_i \langle A|_i\right) \\ j_{i,y} &= \frac{-i}{2}\left(|A\rangle_i \langle C|_i - |C\rangle_i \langle A|_i\right) \\ j_{i,z} &= \frac{1}{2}\left(|C\rangle_i \langle C|_i - |A\rangle_i \langle A|_i\right). \end{aligned} \tag{11.23}$$

These operators, whether or not they describe the true spin of the atom, behave like spin-$1/2$ operators in that $[j_x, j_y] = ij_z/2$ and cyclic permutations. We see that

$$H_{\text{eff},i} = -\alpha_0\left(j_{i,0}S_0(\mathbf{x}_i) + j_{i,z}S_z(\mathbf{x}_i)\right). \tag{11.24}$$

If we assume that the field (or at least the Stokes operator S_z) does not change over the extent of the atoms, we can remove the dependence on $\mathbf{x}_i$. We can then sum to get

$$H_{INT} = -\alpha_0\left(J_0S_0 + J_zS_z\right) \tag{11.25}$$

where $\mathbf{J} = \sum_i \mathbf{j}_i$ describes a spin-N/2 system and obeys

$$[J_a, J_b] = i\varepsilon_{abc}J_cN/2. \tag{11.26}$$

Appendix A

Statistics

This appendix describes some elements of classical statistics and their relation to quantum mechanics.

A.1 Basic concepts

A.1.1 Random variables

A random variable is simply a variable that can take on different values for reasons external to the problem under consideration. For example, if we throw a coin, the "face" on which it lands can be either "heads" or "tails". At least in classical physics, the reason it lands one way or another could in principle be known, a consequence of the initial conditions of the throw such as the velocity, the density of the air through which the coin travels, etc. But if we are not able to or not interested in including such things in our description of the problem, we might as well say the coin will land heads or tails for unknown reasons, and the face we treat as a random variable.

We can put into mathematical notation by writing X (a random variable) for the face, and $X \to \{\text{“}heads\text{”}, \text{“}tails\text{”}\}$ to indicate that X can take on the two values "heads" and "tails." In physics these will usually be the possible "outcomes" of a measurement, and we will use that term in what follows. We might also want to associate the values 0 and 1 to "heads" and "tails", in which case we would write $X \to \{0, 1\}$. $\{0, 1\}$ is the sample space or domain of X. We will use the symbol $\mathcal{X}$ for the domain.

A.1.2 Probabilities and discrete probability distributions

We are of course not only interested in what values a random variable can take, but also how often it takes them, i.e. the probability of getting "heads" or "tails." Probability is a

funny concept: although ordinary people have no trouble with it, mathematicians do. There are different camps: Frequentists and Bayesians, who have different ideas of what probability represents. For physics, this difference in the most essential idea of statistics will be surprisingly non-problematic.

For a "fair coin," i.e. one with equal probability of the two outcomes, we can write the probability as $P(X = \text{“}heads\text{''}) = 1/2$ and $P(X = \text{“}tails\text{''}) = 1/2$ or more compactly as $P(X = x) = 1/2, x \in \{0, 1\}$. Note that we have a different symbol for the random variable X and the outcome x. Physicists often skip this distinction, and might write $P(x) = 1/2, x \in \{0, 1\}$. More generally, a probability P can take on different values for each of the possible outcomes, constrained only by

$$P(X = x) \geq 0 \tag{A.1}$$

$$\sum_{x \in \mathcal{X}} P(X = x) = 1. \tag{A.2}$$

This agrees with the intuitive notions of probability: they are never negative and they sum to one, because *something* should happen. The function P is called the probability distribution (also the probability mass distribution); it describes how the probability is distributed among the different possible outcomes. It is a discrete distribution because X can take on a countable number of values.

A.1.3 Continuous distributions and probability densities

Some random variables can take on continuous values. For example, the arrival time T of a bus might be considered a random variable that can take on a continuum of values, not just discrete values as in the case of a coin. We can't directly apply Eq. (A.2) to T. Rather, we define the probability density function (PDF) $f(x)$, such that

$$f(x) \geq 0 \tag{A.3}$$

$$\int_{x \in \mathcal{X}} f(x)dx = 1. \tag{A.4}$$

The interpretation of f is that $P(X \in [a, b]) = \int_{x=a}^{x=b} f(x)dx$, i.e. the probability to find X between a and b is $\int_{x=a}^{x=b} f(x)dx$. Equivalently, $P(X \in [x, x + dx]) = f(x)dx$. Again, this corresponds to the intuitive ideas of probability.

Another quantity often encountered is the cumulative distribution function

$$F(x) \equiv \int_{x'=-\infty}^{x'=x} f(x')dx' \tag{A.5}$$

which describes $P(X < x)$, the probability for X to be smaller than x. Sometimes F is taken as primary, in which case the PDF is defined as $f(x) \equiv dF(x)/dx$.

A.1.4 Joint probabilities

In many problems we have multiple random variables. A joint probability is simply the probability that a set of random variables give specific outputs. For example, if we have two coins X_1 and X_2, we would write the probability that they are both heads as $P(X_1 = 1, X_2 = 1)$. If these are both fair coins, then the $P(X_1 = 1, X_2 = 1) = 1/4$. This generalizes in the obvious way to more variables.

A.1.5 Marginal probabilities

When we have multiple random variables, but we want to know the distribution of a subset of them, irrespective of what the other variables might do, we are looking for a "marginal distribution." For two discrete variables X_1 and X_2, the marginal distribution for X_1 is given by

$$P(X_1 = x_1) = \sum_{x_2 \in \mathcal{X}_2} P(X_1 = x_1, X_2 = x_2). \tag{A.6}$$

Similarly, for continuous distributions, the joint PDF $f(x_1, x_2)$ describes the probability density to find $X_1 = x_1$ *and* $X_2 = x_2$. If we know $f(x_1, x_2)$ and we want to find the probability density for $X_1 = x_1$ and *any* value of X_2, we need the marginal PDF

$$f(x_1) = \int_{x_2 \in \mathcal{X}_2} f(x_1, x_2) dx_2 \tag{A.7}$$

This generalizes in the obvious way to more variables.

A.1.6 Statistical independence

Two random variables are statistically independent if their joint probability factorizes, i.e. if $P(X_1 = x_1, X_2 = x_2) = P(X_1 = x_1)P(X_2 = x_2)$ for all $x_1 \in \mathcal{X}_1$ and $x_2 \in \mathcal{X}_2$. This generalizes in the obvious way to more variables. It is easy to check that this is consistent with our definition of the marginal distribution:

$$P(X_1 = x_1) = \sum_{x_2 \in \mathcal{X}_2} P(X_1 = x_1, X_2 = x_2) = P_1(X_1 = x_1) \sum_{x_2 \in \mathcal{X}_2} P_2(X_2 = x_2) = P_1(X_1 = x_1) \cdot 1 \tag{A.8}$$

A.1.7 Conditional probabilities

If two variables are not independent, it is interesting to ask how one variable depends on the other. For example, I might as "what is the probability that X_1 is heads, given that I have already see that X_2 is tails?" This is described by the conditional probability distribution $P(X_1 = 1|X_2 = 0)$, where the vertical bar, which could be read "given that," separates the

posterior condition "X_1 is heads" from the prior condition "X_2 is tails." Conditional distributions are defined in relation to joint and and marginal distributions:

$$P(X_1 = x_1, X_2 = x_2) = P(X_1 = x_1|X_2 = x_2)P(X_2 = x_2) \tag{A.9}$$

where $P(X_2 = x_2)$ is the marginal distribution for X_2. The above has the intuitive meaning that the probability of seeing outcomes $X_1 = x_1$ and $X_2 = x_2$ is the probability of seeing $X_2 = x_2$ (under any circumstance) times the probability of seeing $X_1 = x_1$ given that you also see $X_2 = x_2$. A trivial rearrangement gives the conditional probability in terms of the joint probability and the marginal distribution:

$$P(X_1 = x_1|X_2 = x_2) = \frac{P(X_1 = x_1, X_2 = x_2)}{P(X_2 = x_2)} \tag{A.10}$$

A.1.8 Bayes' theorem

Using Eq. (A.9) twice, we can easily derive Bayes' theorem

$$P(X_1 = x_1|X_2 = x_2) = \frac{P(X_2 = x_2|X_1 = x_1)P(X_1 = x_1)}{P(X_2 = x_2)} \tag{A.11}$$

which relates the probability of conditional distributions in two directions, X_1 conditioned on X_2 (on the l.h.s.) to X_2 conditioned on X_1 (on the r.h.s).

A.2 Probability distributions as objects of interest

A lot of physics is concerned with probability distributions. To determine the lifetime of an unstable particle, we ask what time constant τ appears in the exponential distribution $f(t) = \tau^{-1}\exp[-t/\tau]$ that describes the time T at which the particle decays to become something else. T is a random variable. Note that τ is not a random variable: this is a question about a *parameter* in a continuous probability distribution.

We cannot directly measure a parameter in a distribution. We can measure the random variable, e.g. the decay time T in the example above, and try to infer the unknown parameter. This is obviously an indirect method, but it is all that nature allows us, and we need to deal with it if we are going to do physics. Measuring T many times gives us a lot of different values. How do we organize this growing pile of information into something sensible ? A *sample statistic* (or just *statistic*) is a function computed on a bunch of experimental trials to measure a random variable (or a collection thereof). We are already familiar with several statistics: the mean, median, mode, standard deviation, etc.

An *estimator* is a sample statistic that approximates a parameter in the distribution. For example, the sample mean (a statistic) approximates the mean of the distribution (a parameter).

A.2.1 Expectation

The expectation of a discrete random variable X with probability mass distribution $P(X)$ is defined as

$$E[X] \equiv \sum_{x \in \mathcal{X}} P(X = x)x \tag{A.12}$$

which agrees intuitively with the notion of the average or mean of X. For a continuous distribution the expectation is

$$E[X] \equiv \int_{x \in \mathcal{X}} f(x)x \, dx. \tag{A.13}$$

A.2.2 Moments

The n-th moment of X, written μ_n is the expectation of X^n, i.e.

$$\mu_n \equiv E[X^n] \equiv \sum_{x \in \mathcal{X}} P(X = x)x^n \tag{A.14}$$

or, for a continuous distribution,

$$\mu_n \equiv E[X^n] \equiv \int_{x \in \mathcal{X}} f(x)x^n \, dx. \tag{A.15}$$

we see that the first moment, i.e. with $n = 1$, is the mean or expectation.

A.2.3 Variance and standard deviation

The variance is defined as $\text{var}(X) \equiv E[(X - \mu_1)^2]$. It is easy to show that

$$E[(X - \mu_1)^2] = E[X^2 - 2X\mu_1 - \mu_1^2] = E[X^2] - \mu_1^2 = \mu_2 - \mu_1^2 \tag{A.16}$$

Whereas the mean indicates the "centre of mass" of the distribution, the variance indicates in some sense the square of its width. The standard deviation is defined as $\sigma_X \equiv [\text{var}(X)]^{1/2}$.

A.2.4 Covariance

The covariance matrix is the multivariate generalization of the variance. It is a matrix, defined as $\Gamma_{ij} \equiv \text{cov}(X_i, X_j)$, where the covariance between random variables X_i and X_j is

$$\text{cov}(X_i, X_j) \equiv E[X_1 X_2] - E[X_1]E[X_2] \tag{A.17}$$

A.3 Relation to quantum physics

Quantum theory is has a part that is deterministic and a part that is statistical. The evolution of the wave function follows the Schrödinger equation and can in principle be computed exactly from known initial conditions. This is the deterministic part. When we go to measure anything, however, quantum mechanics does not specify what we will observe. Rather, it specifies the probabilities of any possible outcome. That is, quantum mechanics makes predictions about probability distributions for measurement outcomes.

A.3.1 Born rule

The Born rule (due to Max Born) says that: When measuring a discrete observable $\hat{X}$ (the hat indicates this is an operator), with normalized eigenvectors $|\phi_i\rangle$ and corresponding eigenvalues $x_i \in \mathcal{X}$, the probability of seeing measurement outcome x is

$$P(X = x_i) = |\langle x_i|\psi\rangle|^2. \tag{A.18}$$

For a continuous observable X, we have

$$P(X \in [x, x + dx]) \quad = \quad |\langle x|\psi\rangle|^2 \, dx \tag{A.19}$$

or equivalently

$$f(x) \quad = \quad |\langle x|\psi\rangle|^2. \tag{A.20}$$

Note that in the continuous case, the condition $\int f(x)dx = 1$ requires that the wave function $\psi(x) \equiv \langle x|\psi\rangle$ must be normalized such that $\int |\psi(x)|^2 \, dx = 1$. This requires that the eigenvectors are normalized such that $|\langle x'|x\rangle|^2 \propto \delta(x - x')$ where δ indicates the Dirac delta function. This is the normal situation in calculation with wave functions.

A.3.2 Expectation values

In quantum physics we often encounter the "expectation value" of an operator. For a state $|\psi\rangle$ and operator $\hat{X}$ this is $\langle\psi|\hat{X}|\psi\rangle$. The expectation value and the expectation $E[X]$ agree, as a consequence of the Born rule

$$\begin{aligned} E[X] \quad &= \quad \sum_{x_i \in \mathcal{X}} P(X = x_i)x_i = \sum_{x \in \mathcal{X}} |\langle\phi_i|\psi\rangle|^2 x_i = \sum_{x \in \mathcal{X}} \langle\psi|\phi_i\rangle x_i \langle\phi_i|\psi\rangle \\ &= \quad \sum_{x \in \mathcal{X}} \langle\psi|\phi_i\rangle\langle\phi_i|x_i|\psi\rangle = \langle\psi|X|\psi\rangle \end{aligned} \tag{A.21}$$

where we have used the fact that a Hermitian operator can be written as a sum of projectors: $X = \sum_{x \in \mathcal{X}} |\phi_i\rangle\langle\phi_i|x_i$. It is common to omit the state and simply write $\langle X\rangle$ when this will not create confusion.

As with everything involving operators, it is necessary to take care regarding ordering when computing expectations. For example, if X_1 and X_2 are non-commuting operators, i.e. if $X_1X_2 \neq X_2X_1$, then we should not assume $\langle X_1X_2\rangle$ is the same as $\langle X_2X_1\rangle$. For example, for two operators $\hat{X}_1$ and $\hat{X}_2$, the covariance must be defined symmetrically:

$$\mathrm{cov}(\hat{X}_1, \hat{X}_2) \equiv \frac{1}{2}\langle \hat{X}_1\hat{X}_2 + \hat{X}_2\hat{X}_1\rangle - \langle \hat{X}_1\rangle\langle \hat{X}_2\rangle \tag{A.22}$$

to avoid the possibility of a complex-valued covariance, which could arise from the (incorrect) definition $\mathrm{cov}(\hat{X}_1, \hat{X}_2) \equiv \langle \hat{X}_1\hat{X}_2\rangle - \langle \hat{X}_1\rangle\langle \hat{X}_2\rangle$.

A.3.3 Mixed states = statistical mixtures

A mixed state is described by density operator with the form

$$\hat{\rho} = \sum_k w_k \left|\psi_k\right\rangle \left\langle\psi_k\right| \tag{A.23}$$

where w_k are weights that sum to unity $\sum_k w_k = 1$, and $|\psi_k\rangle$ are normalized wave functions (not necessarily orthogonal). The density matrix is useful for describing the average properties of a system that has state $|\psi_k\rangle$ with probability w_k. This is a second way that statistics can enter a quantum problem: as an uncertainty about how the system is prepared. This kind of uncertainty is not specific to quantum physics; it also exists in classical theories, most notably in classical statistical mechanics.

The density matrix represents the density operator in a given basis. If the basis consists of the orthonormal states $\{|\phi_i\rangle\}$, then

$$\rho_{ij} = \left\langle\phi_i\right| \hat{\rho} \left|\phi_i\right\rangle \tag{A.24}$$

$$\hat{\rho} = \sum_{i,j} \rho_{ij} \left|\phi_i\right\rangle \left\langle\phi_i\right| \tag{A.25}$$

A.3.4 Born rule for mixed states

When working with mixed states, it is helpful to have a definition of the Born rule in terms of operators. For an operator X, with eigenvalues $\{x_i\}$ and corresponding eigenvectors $|x_i\rangle$, we can expand the density operator in its eigensystem (this is known as the spectral decomposition of the operator)

$$\hat{X} = \sum_i x_i \left|x_i\right\rangle \left\langle x_i\right| = \sum_i x_i \Pi_{|x_i\rangle} \tag{A.26}$$

where $\Pi_{|x_i\rangle} = |x_i\rangle \langle x_i|$ is a projector onto the state $|x_i\rangle$. We note that

$$P(X = x_i) = \left\langle\psi\right| \Pi_{|x_i\rangle} \left|\psi\right\rangle \tag{A.27}$$

and thus $P(X = x_i)$ is in fact the expectation value of $\Pi_{|x_i\rangle}$. We can take this a step farther by introducing an orthonormal basis $\{|\phi_j\rangle\}$. Because $\sum_j |\phi_j\rangle \langle\phi_j|$ is the identity, we can write

$$P(X = x_i) = \sum_j \langle\psi| \Pi_{|x_i\rangle} |\phi_j\rangle \langle\phi_j| |\psi\rangle = \sum_j \langle\phi_j| |\psi\rangle \langle\psi| \Pi_{|x_i\rangle} |\phi_j\rangle = \mathrm{Tr}[|\psi\rangle \langle\psi| \Pi_{|x_i\rangle}]. \quad \text{(A.28)}$$

Now, combining Equation A.23 with Equation A.28, we discover a very compact expression of the Born rule. If the probability to be in state k is $P(K = k) = w_k$, and given that we are in state k the probability to observe $X = x_i$ is $P(X = x_i|k) = \mathrm{Tr}[|\psi_k\rangle \langle\psi_k| \Pi_{|x_i\rangle}]$, then the probability to see $X = x_i$ regardless of k (the marginal probability), is given by

$$P(X = x_i) \quad = \quad \sum_k P(X = x_i|k)P(K = k) = \sum_k w_k \mathrm{Tr}[|\psi_k\rangle \langle\psi_k| \Pi_{|x_i\rangle}] = \mathrm{Tr}[\hat{\rho}\Pi_{|x_i\rangle}]. \quad \text{(A.29)}$$

This defines a *projective measurement*, also called a von Neumann measurement (even though von Neumann described also other models of measurement!).

A.3.5 Generalized measurements

Not all measurements are projective. Consider for example a measurement on a spin-1/2 particle that can give three answers: "failed to detect anything" (which happens with probability 1/2, regardless of the spin state), "spin up," (which happens with probability 1/2 when the state is $|\uparrow\rangle$), and "spin down" (which happens with probability 1/2 when the state is $|\downarrow\rangle$). This is a possible measurement, but it cannot be described with the mathematics of projective measurement. Any operator on a spin-1/2 particle will have only two eigenstates, and thus any projective measurement on this particle will have at most two possible outcomes. Fortunately, we can easily define a more general measurement model by expanding slightly the Born rule.

Consider a set of operators $\{\hat{A}_i\}$ that are *positive semi-definite*, i.e. for which $\langle\psi| A_i |\psi\rangle \geq 0$ for any $|\psi\rangle$, and that sum to the identity

$$\sum_i \hat{A}_i = \mathbb{1} \quad \text{(A.30)}$$

We can then generalize the Born rule to

$$P(X = x_i) = \mathrm{Tr}[\hat{\rho}\hat{A}_i]. \quad \text{(A.31)}$$

Note that for pure states this is $P(X = x_i) = \langle\psi| A_i |\psi\rangle$. It should now be clear that $P(X = x_i) \geq 0$ and $\sum_i P(X = x_i) = 1$, as required for a probability distribution. This kind of measurement, which includes projective measurement as a special case, is known as a positive operator-valued measure (POVM).

For continuous-valued observables, a continuous POVM describes operators $\{\hat{B}_x\}$, now parametrized by a continuous index x, that obey $\int \hat{B}_x dx = \mathbb{1}$. The PDF for observable X is then given by

$$f(X = x) = \mathrm{Tr}[\hat{\rho}\hat{B}_x]. \quad \text{(A.32)}$$

Appendix B

Quantum theory for quantum optics

This appendix describes some aspects of quantum theory that are either helpful, necessary, or simply illuminating, for the study of quantum optics.

B.1 States vs. Operators

The outcomes of measurements in quantum mechanics are described by expectation values. These could be average quantities such as $\langle x \rangle, \langle x^2 \rangle$, etc. or they could be frequencies of particular outcomes, which are described as the expectation values of a projector. For example a projector onto the state $|\phi\rangle$ is $P_\phi = |\phi\rangle\langle\phi|$, so that the probability of finding a system in state $|\phi\rangle$ is $\langle P_\phi \rangle$. This holds also for continuous-valued probability distributions, such as $|\psi(x)|^2$, which can be found as the expectation values of projectors onto small (in the limit infinitessimal), regions around x.

An expectation value $\langle\phi| A |\phi\rangle$ will evolve as

$$\begin{aligned} \langle A \rangle_\phi (t) &= \langle\phi| U^\dagger(t) A U(t) |\phi\rangle \\ &= \langle\phi(t)|_S A |\phi(t)\rangle_S \\ &= \langle\phi| A_H(t) |\phi\rangle \end{aligned} \tag{B.1}$$

where $U(t)$ is the time-evolution operator, and $|\phi(t)\rangle_S = U(t)|\phi\rangle$ and $A_H(t) = U^\dagger(t) A U(t)$ are the Schrödinger picture state and Heisenberg picture operator, respectively. These pictures are completely equivalent in their results, and in some sense, the difference between them is trivial; what we really want to know is the time evolution $U(t)$. At the same time, the Schrödinger picture is more familiar to most people (it is usually taught first). So why is quantum optics always described in the Heisenberg picture?

One good reason is the classical-quantum correspondence. The equations for the quantum field operators are exactly the same as the equations for the corresponding classical fields, with the consequence that the *average* values of the two theories will agree. Another good reason is that

many problems in field theory are very high-dimensional, so that a full description of the state would be very complicated. In contrast, a description of the few operators that we will eventually measure may be much simpler. Finally, there are a few central problems where operators are easy to work with, and states very difficult. One of these is the beam-splitter, described in an earlier chapter.

B.2 Calculating with operators

B.2.1 Heisenberg equation of motion

Given a Hamiltonian $H(t)$, and an arbitrary operator A, the evolution of A is given by

$$\frac{d}{dt}A = \frac{1}{i\hbar}[A, H(t)] + \partial_t A \tag{B.2}$$

where $\partial_t A$ is the *explicit* time-dependence of A. For example, if $B = q\exp[i\omega t]$ where q is some operator which itself will change in time due to the Hamiltonian, then

$$\frac{d}{dt}B = \frac{1}{i\hbar}[q, H(t)]\exp[i\omega t] + i\omega q\exp[i\omega t]. \tag{B.3}$$

Typically we will avoid operators with explicit time dependence.

B.2.2 Time-dependent perturbation theory

A very useful form of perturbation theory was developed by F. Dyson, based on the interaction picture. The unitary time-evolution operator $U(t)$ obeys the Schrödinger equation[1]

$$i\frac{d}{dt}U(t) = H(t)U(t) \tag{B.4}$$

with the initial condition $U(0) = I$. When the Hamiltonian is broken into two parts as

$$H = H_0 + H', \tag{B.5}$$

it is convenient to also divide U as

$$U = U_0 U_I \tag{B.6}$$

where U_0 is the unperturbed evolution, i.e., $idU_0/dt = H_0 U_0$ so that

$$\begin{aligned} i\frac{d}{dt}(U_0U_I) &= (H_0 + H')U_0U_I \\ i(\frac{d}{dt}U_0)U_I + iU_0(\frac{d}{dt}U_I) &= H_0U_0U_I + H'U_0U_I \\ i\frac{d}{dt}U_I &= U_0^\dagger H' U_0 U_I \\ &\equiv H_I U_I. \end{aligned} \tag{B.7}$$

[1]Note that we are dropping the factors of $\hbar$ here. They are easy to put back, just by noting that the H and $\hbar$ always occur together in the combination $H/\hbar$.

The operator H_I is the interaction picture Hamiltonian, and the evolution U_I describes the change in the state, relative to the state evolution under H_0 alone. For example, $\left\langle U_0^\dagger(t)U(t)\right\rangle_\phi = \langle U_0(-t)U_0(t)U_I(t)\rangle_\phi = \langle U_I(t)\rangle_\phi$ is the overlap of the state $|\phi\rangle$, evolved under H, with the same state evolved under just H_0.

It is easy to check that the evolution of U_I,

$$i\frac{d}{dt}U_I = H_I U_I \tag{B.8}$$

is solved by

$$\begin{aligned} U_I(t) &= \left[1 - i\int_0^t dt'\, H_I(t') - \int_0^t dt' \int_0^{t'} dt''\, H_I(t')H_I(t'') + ...\right] \\ &= \mathcal{T}e^{-i\int_0^t dt'\, H_I(t')}. \end{aligned} \tag{B.9}$$

This expansion is known as the *Dyson series*. Note that the integrals contain the *time-ordered* products $H_I(t')H_I(t'')H_I(t''')...$ with $t' > t'' > t''' > ...$ The symbol $\mathcal{T}$ indicates time-ordering, so that the exponential agrees with the expansion in the line above it. It is also convenient to know $U_I^\dagger$

$$\begin{aligned} U_I^\dagger(t) &= \left[1 + i\int_0^t dt'\, H_I(t') - \int_0^t dt' \int_0^{t'} dt''\, H_I(t'')H_I(t') + ...\right] \\ &= \tilde{\mathcal{T}}e^{+i\int_0^t dt'\, H_I(t')}. \end{aligned} \tag{B.10}$$

Note that here the operators are *anti-time-ordered* ($\tilde{\mathcal{T}}$).

Using these results, in the Schrödinger picture, states evolve as

$$|\psi(t)\rangle = U_0(t)\mathcal{T}e^{-i\int_0^t dt'\, H_I(t')}\,|\psi(t=0)\rangle \tag{B.11}$$

while in the Heisenberg picture operators evolve as

$$\begin{aligned} A(t) &= \tilde{\mathcal{T}}e^{+i\int_0^t dt'\, H_I(t')}U_0^\dagger(t)A(0)U_0(t)\mathcal{T}e^{-i\int_0^t dt'\, H_I(t')} \\ &= \tilde{\mathcal{T}}e^{+i\int_0^t dt'\, H_I(t')}A_0(t)\mathcal{T}e^{-i\int_0^t dt'\, H_I(t')} \end{aligned} \tag{B.12}$$

where $A_0(t) \equiv U_0^\dagger(t)A(0)U_0(t)$ is the unperturbed evolution of the operator A, i.e., the evolution under only H_0 from the initial value $A(0)$. Expanding the first few terms

$$\begin{aligned} A(t) &= A_0(t) \\ &\quad + i\int_0^t dt'\, H_I(t')A_0(t) - iA_0(t)\int_0^t dt'\, H_I(t') \\ &\quad - \int_0^t dt' \int_0^{t'} dt''\, H_I(t'')H_I(t')A_0(t) \\ &\quad - A_0(t)\int_0^t dt' \int_0^{t'} dt''\, H_I(t')H_I(t'') \\ &\quad + \int_0^t dt'\, H_I(t')A_0(t)\int_0^t dt'\, H_I(t') \\ &\quad + O(H_I^3) \end{aligned} \tag{B.13}$$

It is interesting also to take the time derivative of this equation to get the rate of change of A

$$\begin{aligned}\frac{d}{dt}A(t) &= \frac{d}{dt}A_0(t) - i[A_0(t), H_I(t)] \\ &\quad - \int_0^t dt' \left[\left[A_0(t), H_I(t')\right], H_I(t)\right] + O(H_I^3) \end{aligned} \tag{B.14}$$

B.2.3 example: excitation of atoms to second order

We assume a collection of very simple atoms, labeled by the index i, with ground and excited states $|g\rangle_i$ and $|e\rangle_i$, respectively and transition frequencies ω_i. For simplicity, we assume the atoms are all in the same place. We define $b_i \equiv |g\rangle_i \langle e|_i$ and $b_i^\dagger \equiv |e\rangle_i \langle g|_i$ for convenience. If the ground state has zero energy, the Hamiltonian is

$$H = H_{EM} + \sum_i \hbar\omega_i b_i^\dagger b_i + gE^{(+)}b_i^\dagger + g^* E^{(-)} b_i. \tag{B.15}$$

An explanation of this Hamiltonian is given in Chapter **??**. For the moment, just note that the last two terms are reasonable; one term excites an atom while destroying a photon, and the other accomplishes the reverse process. We identify the first two terms as H_0 and the last two as H'. Under H_0 (specifically the part H_{EM}), the field E evolves as $E_0(t)$, the evolution under Maxwell's equations from whatever is the initial condition. Under H_0 the atomic operator b_i evolves as $b_{0,i}(t) = b_i(0)\exp[-i\omega_i t]$. The interaction Hamiltonian is thus

$$H_I = \sum_i g E_0^{(+)}(t) b_i^\dagger(0) e^{i\omega_i t} + g^* E_0^{(-)}(t) b_i(0) e^{-i\omega_i t}. \tag{B.16}$$

We calculate the evolution of n_i as follows: First, $[n_{0,i}, H_0(t)] = 0$, so $n_{0,i}(t)$ is a constant. Also,

$$[n_{0,i}, H_I(t)] = gE_0^{(+)}(t)b_i^\dagger(0)e^{i\omega_i t} - g^* E_0^{(-)}(t) b_i(0) e^{-i\omega_i t} \tag{B.17}$$

and

$$\begin{aligned}\left[\left[n_{0,i}, H_I(t')\right], H_I(t)\right] &= |g|^2 \Big[n_{0,i} E_0^{(+)}(t') E_0^{(-)}(t) e^{-i\omega_i(t-t')} \\ &\quad -(1-n_{0,i}) E_0^{(-)}(t) E_0^{(+)}(t') e^{-i\omega_i(t-t')} \\ &\quad -(1-n_{0,i}) E_0^{(-)}(t') E_0^{(+)}(t) e^{i\omega_i(t-t')} \\ &\quad + n_{0,i} E_0^{(+)}(t) E_0^{(-)}(t') e^{i\omega_i(t-t')} \Big] \end{aligned} \tag{B.18}$$

As a result, the rate of change of n_i is

$$\begin{aligned}\frac{d}{dt}n_i(t) &= -igE_0^{(+)}(t)b_i^\dagger(0)e^{i\omega_i t} + ig^* E_0^{(-)}(t)b_i(0)e^{-i\omega_i t} \\ &\quad -|g|^2\int_0^t dt'\,\Big[n_{0,i}E_0^{(+)}(t')E_0^{(-)}(t)e^{-i\omega_i(t-t')} \\ &\quad -(1-n_{0,i})E_0^{(-)}(t)E_0^{(+)}(t')e^{-i\omega_i(t-t')} \\ &\quad -(1-n_{0,i})E_0^{(-)}(t')E_0^{(+)}(t)e^{i\omega_i(t-t')} \\ &\quad + n_{0,i}E_0^{(+)}(t)E_0^{(-)}(t')e^{i\omega_i(t-t')}\Big] \\ &\quad +O(H_I^3). \end{aligned} \tag{B.19}$$

The first line indicates the conversion of coherence between atomic levels $\langle b\rangle, \left\langle b^\dagger\right\rangle$, to population. The next two lines describe second-order processes. Note that the product $(1-n_{0,i})E_0^{(-)}E_0^{(+)}$ is, very roughly speaking, the probability to find the atom in the ground state times the intensity, indicating that the presence of photons can (in second order) cause the transition $g \to e$. The product $n_{0,i}E_0^{(+)}E_0^{(-)}$ is, again roughly speaking, the probability to find the atom in the excited state times (intensity + vacuum fluctuations). This term produces both spontaneous and stimulated emission $e \to g$.

If an atom starts in its ground state, we have $\langle b_{0,i}\rangle = \left\langle b_{0,i}^\dagger\right\rangle = \langle n_{0,i}\rangle = 0$, so that

$$\frac{d}{dt}\langle n_i(t)\rangle = \frac{|g|^2}{\hbar^2}\int_0^t dt'\,\left\langle E_0^{(-)}(t)E_0^{(+)}(t')e^{-i\omega_i(t-t')} + E_0^{(-)}(t')E_0^{(+)}(t)e^{i\omega_i(t-t')}\right\rangle \tag{B.20}$$

where we have dropped the higher-order terms. A rough interpretation would be: the rate of excitation is the current field $E_0^{(-)}(t)$ times the accumulation of field at the transition frequency $\exp[-i\omega_i t]\int_0^t dt'\,E_0^{(+)}(t')\exp[i\omega_i t']$.

B.2.4 Glauber's broadband detector

Glauber based his celebrated theory of photo-detection on exactly the problem of a collection of atoms excited by a quantum field. Assume the transition frequencies ω_i are distributed over a broad range by some unspecified mechanism of inhomogeneous broadening, so that the detector has a broad bandwidth. Assuming that the detector has some mechanism which produces an electrical output signal in response to excited atoms, we calculate the rate of change of the operator

$$N_e \equiv \sum_i n_i \tag{B.21}$$

which is the total number of excitations. Assuming as above that each atom starts in its ground state, the average rate of detections (increase in number of excited atoms) is

$$\frac{d}{dt}\langle N(t)\rangle = \frac{|g|^2}{\hbar^2}\sum_i\int_0^t dt'\,\left\langle E_0^{(-)}(t)E_0^{(+)}(t')e^{-i\omega_i(t-t')} + E_0^{(-)}(t')E_0^{(+)}(t)e^{i\omega_i(t-t')}\right\rangle. \tag{B.22}$$

Assuming there are many atoms, the sum can be replaced by an integral $\sum_i \to \int d\omega_i \rho(\omega_i)$ where ρ is a "density of states" factor. Assuming that $\rho(\omega) = \rho$ is flat, i.e., broad-band, the integral $\int d\omega_i \rho(\omega_i) \exp[-i\omega_i(t-t')] = 2\pi\rho\delta(t-t')$ and the integral over t' can be evaluated simply. We thus arrive to Glauber's result

$$\frac{d}{dt}\langle N(t)\rangle = 2\pi\rho\frac{|g|^2}{\hbar^2}\left\langle E_0^{(-)}(t)E_0^{(+)}(t)\right\rangle. \tag{B.23}$$

Note that $E_0^{(-)}(t)E_0^{(+)}(t)$ in the RHS is normally-ordered. This implies that in the absence of photons, there will be no detections. This obviously agrees with our experience of photo-detection, and resolves a basic question: why don't photo-detectors (or our eyes!) see the vacuum fluctuations? While there is *energy* in the vacuum, $\left\langle E^2\right\rangle \neq 0$, detectors do not respond to energy, but rather to photons.

B.3 Second quantization

In Chapter 2, we used canonical quantization to find a quantum theory of the electro-magnetic field. This theory describes the evolution of field operators, which can be expressed in terms of creation and annihilation operators for field modes, and these behave like quantum mechanical harmonic oscillators. In this description, a photon is an excitation of a harmonic oscillator.

This theory may not *seem* to have much in common with the quantum mechanics of other particle such as electrons or atoms, but it does. We can describe other bosonic particles in a very similar, field-theoretic way. This gives a lot of insight into what is going on with light. This is the subject of *second quantization*, the description of collections of particles using field operators. The name "second quantization" is a bit misleading. In the quantization of the EM field, we invented a new quantum theory based on an old, classical theory. In second quantization there is no new theory, it is simply a way to write an old theory (quantum mechanics) using field-theoretic notation. When we are done, quantum mechanics will look just like quantum optics.

Collections of bosonic particles, such as photons, ^{4}He, and most of the atoms that can be laser cooled, are described by exchange-symmetric wave-functions. For the moment, we work with time-independent wave-functions. Later we will consider how the states evolve. If we write two orthonormal single-particle wave-functions as $\phi_1(x), \phi_2(x)$, then possible two-particle wave-functions are

$$\begin{aligned}\Psi_{\phi_1\phi_2}(x_1,x_2) &= \frac{1}{\sqrt{2}}\left[\phi_1(x_1)\phi_2(x_2)+\phi_1(x_2)\phi_2(x_1)\right]\\ \Psi_{\phi_1^2}(x_1,x_2) &= \frac{1}{2}\left[\phi_1(x_1)\phi_1(x_2)+\phi_1(x_2)\phi_1(x_1)\right]=\phi_1(x_1)\phi_1(x_2).\end{aligned} \tag{B.24}$$

We note that both are symmetric under the exchange $x_1 \leftrightarrow x_2$. Because of the exchange symmetry requirement, Ψ is completely determined by the wave-functions $\phi_1(x), \phi_2(x)$, and by how many particles occupy each single-particle state. For this reason, we can use the labels $\phi_1\phi_2$ and ϕ_1^2 to describe these two states, where the superscript indicates the occupancy.

A general N-particle bosonic state is described by

$$\Psi_{\phi_1^{n_1}\phi_2^{n_2}\ldots}(x_1, x_2 \ldots) \equiv \frac{1}{\sqrt{N! n_1! n_2! \ldots}} \sum_{\mathcal{P}\{x_1, x_2, \ldots, x_N\}} \overbrace{\phi_1(x_1)\phi_1(x_2)\ldots\phi_1(x_{n_1})}^{n_1}\overbrace{\phi_2(x_{n_1+1})\ldots}^{n_2} \quad \text{(B.25)}$$

where the sum is over all possible permutations of $\{x_1, x_2, \ldots, x_N\}$ and the factor under the square root is needed to preserve normalization. We will write the state which has this wave function as $|\phi_1^{n_1}\phi_2^{n_2}\ldots\rangle$.

If our goal were to treat a system with a fixed number of particles, we could work with the wave-function above, for example write down the Schrödinger equation that it obeys, try to find solutions, etc. But there are many situations where the number of particles is not fixed. Examples include reactions such as $\gamma \rightarrow e^- + e^+$ in high energy physics, or absorption of photons by material, which in a semiconductor might produce the reaction $\gamma \rightarrow e^- + h^+$ where h^+ indicates a "hole," a positive charge carrier. If we want to think about these situations, we need to be able to describe states with variable numbers of photons. For example, we could write a state

$$|\Sigma\rangle = c_0 |0\rangle + c_1 |\phi_a\rangle + c_2 |\phi_b\phi_c\rangle + \ldots \quad \text{(B.26)}$$

where the c_1 term describes the part of the state with one particle, the c_2 part describes the part of the state with two particles, etc. The c_0 part of the state is something new, and clearly necessary if particles can be destroyed. It describes the amplitude for having *no* particles, which we call "vacuum" and write as $|0\rangle$.

We also see that now we can describe operators which change the number of particles, for example $|\phi_2\rangle\langle\phi_1\phi_2|$ is a valid operator. It converts the state $|\phi_1\phi_2\rangle$ into the state $|\phi_2\rangle$ (and annihilates anything else). This operator would only have a nonzero expectation value if the state contained both $|\phi_1\phi_2\rangle$ and $|\phi_2\rangle$, i.e., if the state had an indeterminate number of particles. We can define a_{ϕ_i}, the annihilation operator for the state ϕ_i through

$$a_{\phi_i} \left|\phi_1^{n_1}\phi_2^{n_2}\ldots\phi_i^{n_i}\ldots\right\rangle = \sqrt{n_i} \left|\phi_1^{n_1}\phi_2^{n_2}\ldots\phi_i^{n_i-1}\ldots\right\rangle. \quad \text{(B.27)}$$

This describes the removal of a particle from the single-particle state ϕ_i. The factor $\sqrt{n_i}$ will be the subject of a problem. The creation operator $a_{\phi_i}^\dagger$ for the same state acts as

$$a_{\phi_i}^\dagger \left|\phi_1^{n_1}\phi_2^{n_2}\ldots\phi_i^{n_i}\ldots\right\rangle = \sqrt{n_i+1} \left|\phi_1^{n_1}\phi_2^{n_2}\ldots\phi_i^{n_i+1}\ldots\right\rangle. \quad \text{(B.28)}$$

It is easy to check that

$$|\phi_1^{n_1}\phi_2^{n_2}\ldots\rangle = \left[\frac{(a_{\phi_1}^\dagger)^{n_1}}{\sqrt{n_1!}}\frac{(a_{\phi_2}^\dagger)^{n_2}}{\sqrt{n_2!}}\ldots\right]|0\rangle \quad \text{(B.29)}$$

and that the creation and annihilation operators obey

$$\begin{aligned}[a_{\phi_i}, a_{\phi_j}] &= [a_{\phi_i}^\dagger, a_{\phi_j}^\dagger] = 0 \\ [a_{\phi_i}, a_{\phi_j}^\dagger] &= \delta_{ij}. \end{aligned} \quad \text{(B.30)}$$

Clearly there is a similarity between these operators and the creation and annihilation operators for the modes of the electromagnetic field.

Field operators

The crucial step in second quantization (some might say the only step) is the introduction of *field operators*. We consider the case where the $\phi_i(x)$ are the single-particle energy eigenstates with energies $\hbar\omega_i$. A single particle wave-function evolves as

$$\phi_i(x,t) = e^{-i\omega_i t}\phi_i(x,0). \tag{B.31}$$

We also assume for simplicity that the particles are non-interacting, so that the evolution of a multi-particle wave function is

$$\phi_1(x,t)\phi_2(x,t)\ldots = e^{-i\omega_1 t}\phi_1(x,0)e^{-i\omega_2 t}\phi_2(x,0)\ldots \tag{B.32}$$

We define the field operator $\Phi(x,t)$ as

$$\Phi(x,t) \equiv \sum_i a_{\phi_i}\phi_i(x,t) = \sum_i a_{\phi_i}\phi_i(x,0)e^{-i\omega_i t} \tag{B.33}$$

with Hermitian conjugate

$$\Phi^\dagger(x,t) \equiv \sum_i a^\dagger_{\phi_i}\phi^*_i(x,t) = \sum_i a^\dagger_{\phi_i}\phi^*_i(x,0)e^{i\omega_i t}. \tag{B.34}$$

These field operators are very powerful, and have a simple interpretation. Consider the state $\Phi^\dagger(x,0)\,|0\rangle$, which clearly contains only one particle. The state is

$$|\psi\rangle = \sum_i \phi^*_i(x,0)\,|\phi_i\rangle \tag{B.35}$$

with wave-function

$$\Psi(x_1) = \sum_i \phi^*_i(x,0)\phi_i(x_1,0) = \delta(x_1 - x) \tag{B.36}$$

by completeness of the states ϕ. From this we see that $\Phi^\dagger(x)$ creates a particle at position x. Similarly, $\Phi(x)$ destroys a particle at position x.

We note that Φ and $\Phi^\dagger$ are very similar to the positive and negative frequency parts of the quantized electric field $\mathbf{E}^{(+)}$ and $\mathbf{E}^{(-)}$ defined in Chapter 4.

$$\begin{aligned}
\hat{\mathbf{E}}^{(+)}(\mathbf{r},t) &= i\sum_{k,\alpha}\sqrt{\frac{\hbar\omega_k}{2\varepsilon_0}}\hat{a}_{k,\alpha}\mathbf{u}_{k,\alpha}(\mathbf{r})e^{-i\omega_k t} \\
\hat{\mathbf{E}}^{(-)}(\mathbf{r},t) &= -i\sum_{k,\alpha}\sqrt{\frac{\hbar\omega_k}{2\varepsilon_0}}\hat{a}_{k,\alpha}\mathbf{u}^*_{k,\alpha}(\mathbf{r})e^{i\omega_k t}
\end{aligned} \tag{B.37}$$

At this point, we have turned ordinary quantum mechanics of many bosons into a field theory. Perhaps surprisingly, the same can be done for fermions, which have exchange anti-symmetric wave functions and thus obey the Pauli exclusion principle. The only change that is necessary is

to replace the creation and annihilation operators a and $a^\dagger$ for creation and annihilation operators b and $b^\dagger$ which obey the anti-commutation relation $\{b_i, b_j^\dagger\} = \delta_{ij}$.

You might ask, why would we want ordinary quantum mechanics to look like field theory? Because the tools of field theory can be applied very easily to situations with variable numbers of particles. Examples would be systems in contact with a reservoir, thermal production of particles such as phonons, and Bose-Einstein condensation. The field operators often behave something like system-wide versions of a single-particle wave function. For example,

$$\int dx\, \Phi^\dagger(x)\Phi(x) = \int dx \sum_{ij} \phi_i^*(x)\phi_j(x) a_{\phi_i}^\dagger a_{\phi_j} = \sum_i a_{\phi_i}^\dagger a_{\phi_i} = \sum_i n_i, \tag{B.38}$$

where the middle step follows from the orthonormality of the states ϕ. Since the last expression is the total number of particles, we conclude that $\Phi^\dagger(x)\Phi(x)$ is a sort of particle-density operator. Also, any single-particle operator A, with matrix elements

$$A_{ij} \equiv \langle \phi_i | \, A \, | \phi_j \rangle = \int dx\, \phi_i^*(x) A \phi_j(x) \tag{B.39}$$

will have a multi-particle equivalent. Explicitly,

$$\int dx\, \Phi^\dagger(x) A \Phi(x) = \int dx \sum_{ij} \phi_i^*(x) A \phi_j(x) a_{\phi_i}^\dagger a_{\phi_j} = \sum_{ij} a_{\phi_i}^\dagger A_{ij} a_{\phi_j}. \tag{B.40}$$

This describes, in a very natural way, the total value of A, including the contributions of all particles. Similarly, a two-particle interaction potential $V(x, x')$ can be applied to the whole collection of particles with an operator $1/2 \int dx\, dx'\, \Phi^\dagger(x)\Phi^\dagger(x')V(x, x')\Phi^\dagger(x')\Phi^\dagger(x)$.