Lecture Notes on

Atomic and Optical Physics

Kevin Zhou

kzhou7@gmail.com

These notes cover atomic and optical physics, with a focus on quantum optics. Some intuition from quantum field theory is assumed. The primary sources were:

- Robert Littlejon's Physics 221 notes. This is an exceptional set of graduate-level quantum mechanics notes, which covers some atomic physics at the same high standard of quality. It doesn't cover much of it, but what it does cover is made much clearer than in any book.
- MIT's 8.421 lecture notes. A set of notes with an emphasis on radiative transitions and useful intuition. But beware: occasionally the notes confidently make assertions that are false.
- Foot, *Atomic Physics*. An accessible book on the interaction of quantum atoms with classical light. Devotes considerable attention to atomic energy levels. Its dense typography makes it somewhat exhausting to read.
- Fox, *Quantum Optics*. This book is by far the most gentle introductory quantum optics book. It does a good job of covering the history of the subject, as well as the ideas behind important experiments. However, many important theoretical details are skimmed over, and very little formalism is used; even some parts of first-semester quantum mechanics are not assumed. Depending on your disposition, you may find this book either refreshing or frustratingly vague.
- Gerry and Knight, *Introductory Quantum Optics*. Another nice and accessible book, with a more theoretical focus, written in a clean style and assuming undergraduate quantum mechanics.
- Schlosshauer, *Decoherence and the Quantum-to-Classical Transition*. A clear, if somewhat wordy book on the essentials of decoherence, and its philosophical implications.

The most recent version is here; please report any errors found to kzhou7@gmail.com.

Contents

1 Semiclassical Radiative Transitions **1**
- 1.1 Transition Rates . . . 1
- 1.2 Lineshapes . . . 6
- 1.3 Lasers . . . 7

2 Quantum Optics **10**
- 2.1 Background . . . 10
- 2.2 Photon Statistics . . . 14
- 2.3 Correlation Functions . . . 18
- 2.4 Quantum Modes . . . 22
- 2.5 Quantum Correlation Functions . . . 25

3 Quantum States of Light **34**
- 3.1 Phase Space Distributions . . . 34
- 3.2 Squeezed States . . . 38
- 3.3 Other Nonclassical States . . . 39

4 Radiative Transitions **40**
- 4.1 Rabi Oscillations . . . 40
- 4.2 The Jaynes–Cummings Model . . . 45
- 4.3 Atoms in Cavities . . . 48

5 Open Systems **51**
- 5.1 Decoherence . . . 51
- 5.2 Models of Decoherence . . . 55
- 5.3 Master Equations . . . 56
- 5.4 Quantum Monte Carlo . . . 56
- 5.5 Quantum Interpretations . . . 56

6 Precision Atomic Experiments **57**
- 6.1 Cold Atoms . . . 57
- 6.2 Atom Interferometry . . . 61
- 6.3 Ion Traps . . . 64
- 6.4 Atomic Clocks . . . 64
- 6.5 Rydberg Atoms . . . 64

1 Semiclassical Radiative Transitions

1.1 Transition Rates

In this section, we cover radiative transitions without using a full quantum formalism.

- We consider a two-level system with energies E_1 and E_2, which can absorb or emit light with frequency $\hbar\omega = E_2 - E_1$. This can be described phenomenologically by the Einstein coefficients.
- Letting N_1 and N_2 be the populations for the states, we have

  $$\frac{dN_1}{dt} = -B_{12}N_1u(\omega), \quad \frac{dN_2}{dt} = -A_{21}N_2 - B_{21}N_2u(\omega).$$

  Here, the three terms represent absorption, spontaneous emission, and stimulated emission. For the first and third, we have pulled out factors of $u(\omega)$, the spectral energy density of the electromagnetic field evaluated at ω.
- Both absorption and stimulated emission are intuitive classically. If a simple harmonic oscillator is driven with a sinusoidal force, then it can either gain or lose energy as a result, at a rate proportional to the driving force. The new and puzzling phenomenon is spontaneous emission, which is sometimes loosely described as "simulated emission by quantum fluctuations".
- Einstein famously was able to determine A_{21} from the others using pure thermodynamics, without needing a complete quantum theory of light. We use the fact that for a population of atoms in thermal equilibrium with radiation of temperature T, we must have

  $$\frac{N_2}{N_1} = \frac{g_2}{g_1}\exp\left(-\frac{\hbar\omega}{k_BT}\right)$$

  where we have included state degeneracies g_i for generality.
- Furthermore, Planck's law gives that

  $$u(\omega) = \frac{\hbar\omega^3}{\pi^2c^3}\frac{1}{e^{\hbar\omega/k_BT}-1}.$$

  By demanding that all these equations be consistent for all temperatures, we find

  $$g_1B_{12} = g_2B_{21}, \quad A_{21} = \frac{\hbar\omega^3}{\pi^2c^3}B_{21}.$$

- We can then compute the Einstein coefficients using time-dependent perturbation theory. Fermi's golden rule states that the transition rate is

  $$W = \frac{2\pi}{\hbar}|M|^2g(\hbar\omega)$$

  where M is the matrix element for the transition, and $g(\hbar\omega)$ is the density of final states.
- Now we run into a subtlety. Fermi's golden rule only makes sense when the final states form a continuum, so we can talk about a density of states. In fact, if the final states are not a continuum, then a constant "transition rate" doesn't even make sense. For example, for driving to a discrete state, the amplitude grows initially linearly, meaning a quadratic probability and

a linearly growing transition rate. For precisely this reason, the only example of absorption we computed in the notes on Undergraduate Physics was the photoelectric effect, where the electron is ejected and thus enters a continuum of states. Another example would be a transition between two continuous electron energy bands in solid state physics.

- In order to apply Fermi's golden rule in this context, we need to think about the continuum of photon states, which requires bringing in the theory of quantum optics we were purportedly avoiding. In this case, it's easiest to think about spontaneous emission, as the final states are just one-photon states. We can write down their density of states using the usual heuristic that a state occupies a volume h^3 of phase space, times 2 for polarizations.

- Above, we have assumed the atom is in free space. The density of states can be modified significantly by changing the environment of the atom. In the context of cavity QED, this is called the Purcell effect.

- Now we discuss the form of the matrix element. We will go through the route of quantum electrodynamics, before reducing it to a undergraduate semiclassical problem. This is the logical route, though the opposite of the historic route.

- The Hamiltonian for the quantum electromagnetic field, along with particles labeled by α, is

$$H = \sum_i \hbar\omega_i a_i^\dagger a_i + \sum_\alpha \frac{1}{2m}\left(\mathbf{p}_\alpha - \frac{q_\alpha}{c}\mathbf{A}(\mathbf{r}_\alpha)\right)^2 + \sum_{\alpha<\beta}\frac{q_\alpha q_\beta}{|\mathbf{r}_\alpha - \mathbf{r}_\beta|} - \sum_\alpha \boldsymbol{\mu}_\alpha \cdot \mathbf{B}(\mathbf{r}_\alpha).$$

  This leaves out a number of physical effects, such as particle creation, or retardation effects for the electrostatic interaction of the particles with each other, but it will be sufficient for our purposes.

- Restricting to a single atom, the matter Hamiltonian is

$$H_{\text{matter}} = \frac{1}{2m}\left(\mathbf{p} + \frac{e}{c}\mathbf{A}(\mathbf{r})\right)^2 + U(r) + \frac{e}{mc}\mathbf{S}\cdot\mathbf{B}(\mathbf{r}).$$

  If the electric fields associated with the light are small compared to the atomic fields, which holds in most practical situations, then the A^2 term is negligible, so we have perturbing Hamiltonian

$$H_1 = \frac{e}{mc}\left(\mathbf{p}\cdot\mathbf{A}(\mathbf{r}) + \mathbf{S}\cdot\mathbf{B}(\mathbf{r})\right).$$

- We wish to calculate the matrix element $\langle f|H_1|i\rangle$. For spontaneous emission, we begin with zero photons, so the fields in H_1 simply act on the photon vacuum to create a state with one photon. This gives a factor of the vacuum photon field amplitude,

$$A \sim \sqrt{\frac{2\pi\hbar c^2}{\omega V}}.$$

  This normalization comes from the canonical commutation relations, but it can also be heuristically derived by noting that $\epsilon E^2 V/2 \sim \hbar\omega/4$ for the vacuum state, and $E \sim (\omega/c)A$.

- Combining these factors, we have

$$\langle f|H_1|i\rangle = \frac{e}{mc}\sqrt{\frac{2\pi\hbar c^2}{V\omega_k}}\, i\hbar M^*, \quad M = \frac{i}{\hbar}\langle 2|(\mathbf{p}\cdot\boldsymbol{\epsilon}_\lambda + i\mathbf{S}\cdot(\mathbf{k}\times\boldsymbol{\epsilon}_\lambda))e^{i\mathbf{k}\cdot\mathbf{r}}|1\rangle$$

  where $|1\rangle$ and $|2\rangle$ are atomic states, $\boldsymbol{\epsilon}$ is a polarization, and $\mathbf{k}$ is the wavevector of the produced photon. Note that $\mathbf{p}\cdot\boldsymbol{\epsilon}_\lambda$ commutes with $e^{i\mathbf{k}\cdot\mathbf{r}}$ because of the transversality condition $\boldsymbol{\epsilon}\cdot\mathbf{k} = 0$.

- We must then average over final states according to the direction of $\mathbf{k}$, and then multiply by the density of states, a calculation we omit, but describe parametrically below.
- If we were describing absorption or stimulated emission, the computation would be slightly more complicated because we would not start in the vacuum state. Emitting into a state that already has n photons yields a factor of $\sqrt{n+1}$ in the amplitude, while absorbing from a state with n photons comes with a factor of $\sqrt{n}$. (In the semiclassical limit, these factors both become $\sqrt{n}$, which makes the amplitude proportional to the electromagnetic field amplitude, as we would expect from thinking in terms of classical forces.)
- Another complication in this case is that the electromagnetic field might be thermal, requiring an additional probabilistic sum. In this case, the relations between Einstein coefficients ensure that in equilibrium, detailed balance holds for each reaction, which in this context follows from the Hermiticity of H_1.
- A semiclassical theory of radiation can actually yield many of the conclusions we will draw below. The point is that absorption can be treated semiclassically. Then spontaneous emission, which makes no sense classically, can be inferred from an Einstein relation.

Now we investigate the atomic matrix element M.

- We note that in atomic units, the size of the atom is $a \sim 1/Z$, while the frequency of the emitted radiation is $\omega \sim Z^2$. Since $c \sim 1/\alpha$ in atomic units, we have

  $$\mathbf{k} \cdot \mathbf{r} \sim Z\alpha$$

  which is small for light atoms. Furthermore, the term $i\mathbf{S} \cdot (\mathbf{k} \times \boldsymbol{\epsilon}_\lambda)$ is a factor of $Z\alpha$ smaller than $\mathbf{p} \cdot \boldsymbol{\epsilon}_\lambda$.
- We can hence expand in a series in $Z\alpha$. For example, at the first two orders,

  $$M^{(0)} = \frac{i}{\hbar}\langle 2|\boldsymbol{\epsilon}_\lambda \cdot \mathbf{p}|1\rangle, \quad M^{(1)} = -\frac{1}{\hbar}\langle 2|\mathbf{S} \cdot (\mathbf{k} \times \boldsymbol{\epsilon}_\lambda) + (\boldsymbol{\epsilon}_\lambda \cdot \mathbf{p})(\mathbf{k} \cdot \mathbf{r})|1\rangle.$$

  The zeroth order term can be recast in a more familiar form. We note that

  $$[\mathbf{r}, H_0] = i\hbar \frac{\mathbf{p}}{m}.$$

  Plugging this in, we have

  $$M^{(0)} = \frac{m}{\hbar^2}\boldsymbol{\epsilon}_\lambda \cdot \langle 2|\mathbf{r}H_0 - H_0\mathbf{r}|1\rangle = -\frac{m\omega}{\hbar}\boldsymbol{\epsilon}_\lambda \cdot \langle 2|\mathbf{r}|1\rangle.$$

  The matrix element is thus proportional to the matrix element of the electric dipole operator. The transitions resulting from this term are hence called electric dipole, or E1 interactions.
- The next term may be expanded to yield matrix elements of the magnetic dipole moment $\mathbf{L} + 2\mathbf{S}$ and the electric quadrupole moment Q_{ij}. The resulting transitions are called M1 and E2 transitions. These are a factor of $Z\alpha$ smaller in amplitude, and hence $(Z\alpha)^2$ smaller in rate.
- Note that we didn't encounter higher moments in the notes on Undergraduate Physics because there we were primarily concerned with constant fields, where only the E1 and M1 terms exist. When we analyzed the photoelectric effect, we didn't expand $e^{i\mathbf{k}\cdot\mathbf{x}}$ at all, because we were dealing with X-ray photons, where $\mathbf{k}$ was large.

- The same holds for nuclear transitions, where $\mathbf{k} \cdot \mathbf{x}$ is not small, and it is better to write the photon states in terms of eigenstates of (π, J^2, J_z) known as vector spherical harmonics. These are similar to ordinary spherical harmonics, with some twists because of the transverse nature of electromagnetic waves; for example, there is no analogue of the s-wave.

- The E1 term is the simplest and probably the easiest to guess, without the benefit of foresight. In fact, using the Thomas–Reiche–Kuhn sum rule, one can show that the rate of energy absorption due to E1 transitions simply matches what one would expect for classical "electron on a spring" dipoles. The dipole moment must be computed with quantum mechanics, but the resulting $\hbar$ cancels out with the $\hbar$ in the energy per photon. We have now returned all the way to classical physics, and indeed historically, this was one of the first classical checks of a quantum result.

- For example, in undergraduate physics one learns that due to classical radiation, an atom would completely collapse in roughly nanoseconds. Since the power matches with that of electric dipole radiation in quantum mechanics, we learn that states that can decay by such transitions have lifetimes of order nanoseconds. Quantum mechanics, then, additionally explains why the final state can be stable.

- In atomic systems, electric dipole transitions are called allowed, and all others are called forbidden, since they are so much slower. Typically, if there exists a path from one state to another by only electric dipole transitions, that path will dominate all others.

- In an electric dipole transition for hydrogen, the Wigner–Eckart theorem gives

$$|\Delta j| \le 1, \quad |\Delta m_j| \le 1, \quad \pi_f = -\pi_i.$$

  In fact, we can say more. For the simple model of hydrogen, we know the transition does not affect the spin degree of freedom. Hence we have the more restrictive rules

$$|\Delta \ell| = 1, \quad |\Delta m| \le 1, \quad \Delta s = \Delta m_s = 0.$$

  Also note that by representation theory, $j = 0 \to 0$ is forbidden, because $0 \notin 0 \otimes 1$.

- Now consider a general, many-electron atom with total angular momentum J. By the same rigorous arguments, we have

$$|\Delta J| \le 1, \quad |\Delta M_J| \le 1, \quad \pi_f = -\pi_i$$

  and additionally as before, $J = 0 \to 0$ is forbidden.

- However, in general it is difficult to separate the spin and orbital contributions, because the spin-orbit interaction mixes them. In cases where this is negligible (i.e. when LS-coupling applies), and hence L and S have definite values, we can conclude

$$|\Delta L| = 1, \quad \Delta S = 0$$

  where parity excludes $\Delta L = 0$. Furthermore, precisely one electron in the electron configuration changes its state.

- In a magnetic dipole transition, we rigorously have

$$|\Delta J| \leq 1, \quad |\Delta M_J| \leq 1, \quad \pi_f = \pi_i.$$

  While the electric dipole operator is purely spatial, the magnetic dipole operator isn't, and hence can flip the spin. In fact, since it mixes orbital and spin contributions in the same way that spin-orbit coupling does, it usually is the dominant transition mechanism between different states in the same LS multiplet, as parity forbids $E1$ transitions.

- Again, when LS-coupling holds, we also have further selection rules. For instance, we have $|\Delta L| \leq 1$ again, and combining this with parity gives $|\Delta L| = 0$.

- Finally, consider electric quadrupole transitions. We rigorously have

$$|\Delta J| \leq 2, \quad |\Delta M_J| \leq 2, \quad \pi_f = -\pi_i$$

  while in addition, $J = 0 \to 1$, $J = 0 \to 1$, $J = 1 \to 0$, and $J = 1/2 \to 1/2$ are all forbidden. Since the electric quadrupole operator is purely spatial, when LS-coupling holds,

$$|\Delta L| = 0, 2, \quad \Delta S = 0$$

  where we used parity to eliminate $|\Delta L| = 1$.

- In some extreme cases, such as $J = 0 \to 0$, *all* single-photon radiative transitions are forbidden; such states are called metastable. The atom can only lose its energy by collision with other atoms, or by rare multi-photon processes, which occur at higher order in perturbation theory.

Example. Estimating spontaneous emission rates in atomic units. We recall that in atomic units,

$$\hbar = m = e = 1, \quad c = \frac{1}{\alpha} \approx 137.$$

In these units, the Bohr radius, the Rydberg, and the semiclassical velocity and frequency of the electron in hydrogen are all equal to 1. For an E1 transition, the rate is

$$\Gamma \sim \left(\frac{e^2}{m^2c^2}\right)\left(\frac{c^2}{\omega V}\right)\left(\frac{V\omega^2}{c^3}\right)|\mathcal{M}^{(0)}|^2 \sim \frac{\omega e^2}{c^3m^2}|\mathcal{M}^{(0)}|^2 \sim \alpha^3\omega|\mathcal{M}^{(0)}|^2$$

where the factors are the interaction strength, the photon field normalization, the density of states of the photon field, and the atomic matrix element. For atomic number Z, the energy levels are of order $\omega \sim Z^2$. The matrix element involves a momentum, which goes as Z. Then

$$\Gamma_{E1} \sim \alpha^3 Z^4 \sim Z^4(1\,\text{ns})^{-1}.$$

Thus even the fastest transitions in hydrogen occur over the span of $1/\alpha^3 \sim 10^6$ electron orbits. We know M1 and E2 transitions have matrix elements penalized by αZ, so for them, $\Gamma \sim \alpha^5 Z^6$.

Example. Consider the 21 cm line in hydrogen. This decay proceeds by an M1 transition, but the splitting is very small since it is due to hyperfine structure, $\omega \sim \alpha^2(m_e/m_p)$. Then we have

$$\Gamma \sim \alpha^3\left(\alpha^2\frac{m_e}{m_p}\right)\alpha^2 \sim \alpha^7\frac{m_e}{m_p}.$$

(this is far too large, fix)

Example. The decay $2s_{1/2} \to 1s_{1/2}$. This cannot proceed as an E1 transition by parity. Meanwhile, since the $2s$ and $1s$ orbitals are nearly orthogonal, the M1 transition is strongly suppressed, i.e. it occurs at higher order in α than one would naively expect. Finally, E2 transitions are forbidden since $J = 1/2 \to 1/2$ is, with the same logic holding for all higher multipole moments. As a result, the dominant decay route is actually two-photon emission, using the electric dipole term twice.

In second-order perturbation theory, one needs to sum over all intermediate states. This suggests that the dominant contribution to two-photon emission is the $2p_{1/2}$ state, because it is intermediate in energy between $2s_{1/2}$ and $1s_{1/2}$ by the Lamb shift, and hence the emission can proceed on-shell. However, this is not the dominant contribution because it has a tiny phase space to occur, since the $2p_{1/2}$ and $2s_{1/2}$ states are so close in energy. **(make more quantitative)**

1.2 Lineshapes

In this section we investigate lineshapes, the frequency distribution of radiation emitted in atomic transitions.

- The shape of a spectral line is described by a spectral lineshape function $g(\omega)$, which peaks at the line center $\Delta E/\hbar$, and whose integral is normalized to one. A standard measure for its width is the full width at half maximum (FWHM) $\Delta\omega$.

- Broadening mechanisms can be homogeneous or inhomogeneous. In the former case, all of the individual atoms behave in the same way. In the latter, the atoms behave differently and contribute to different parts of the spectrum. Generally, homogeneous mechanisms give Lorentzian lineshapes and inhomogeneous ones give Gaussian lineshapes.

- The natural/radiative broadening is due to the finite lifetime of an excited state, and is required by the energy-time uncertainty principle. For an exponential decay with time constant τ, one gets a Lorentzian with FWHM $1/\tau$,

$$g(\omega) = \frac{\Delta\omega}{2\pi} \frac{1}{(\omega - \omega_0)^2 + (\Delta\omega/2)^2}, \quad \Delta\omega = \frac{1}{\tau}.$$

  Semiclassically, this holds because a Lorentzian is the Fourier transform of a decaying oscillation. More rigorously, it can be derived quantum mechanically, where the Lorentzian comes from propagator factors/energy denominators.

- If the time between atomic collisions is less than the lifetime, then there is additional collisional/pressure broadening. The parameter τ above is replaced with the mean collision time

$$\tau \sim \frac{1}{\sigma P} \left(\frac{\pi m k_B T}{8} \right)^{1/2}.$$

  This is significant at standard temperatures and pressures, which is why "low pressure" discharge lamps are used in spectroscopy.

- Doppler broadening comes from the Doppler shift from the longitudinal velocity component of the atoms, given by the Maxwell-Boltzmann distribution,

$$p(v_\parallel) \propto \exp\left(-\frac{m v_\parallel^2}{2 k_B T} \right).$$

This yields a Gaussian lineshape,

$$g(\omega) = \frac{c}{\omega_0}\sqrt{\frac{m}{2\pi k_B T}}\exp\left(-\frac{mc^2(\omega-\omega_0)^2}{2k_B T\omega_0^2}\right)$$

whose FWHM is

$$\Delta\omega = 2\omega_0\left(\frac{(2\log 2)k_B T}{mc^2}\right)^{1/2}.$$

For a gas at standard temperature, this is a few orders of magnitude larger than the natural width, so it dominates for gases at low pressure.

- In solids, pressure and Doppler broadening are not relevant. However, atoms can de-excite by non-radiative transitions, such as by phonon emission. This reduces the lifetime, leading to a broadened Lorentzian lineshape just like pressure broadening.
- Inhomogeneities of the host medium can also cause environmental broadening, another type of inhomogeneous broadening.
- In semiconductors, absorption and emission work a bit differently, in interband transitions. Emission is sharply peaked at the band gap, while absorption occurs at or above the band gap. This is because emission occurs due to recombination of electrons and holes. Before the main emission occurs, the electrons and holes rapidly lose energy within their bands, leaving only the band gap energy available for emission. The width of the emission line is determined by thermal or inhomogeneity effects.

Note. The quality factor of a transition is its central frequency divided by its FWHM. A typical optical excitation has frequency of order $10^{15}\,\mathrm{Hz}$. At room temperature, Doppler broadening yields $Q \sim 10^6$. If we remove Doppler broadening, by confining the atom in a lattice, and use a metastable level with a lifetime of about one second, we get $\Delta f \sim 1\,\mathrm{Hz}$, so $Q \sim 10^{15}$.

By contrast, a typical mechanical oscillator, such as one in a quartz clock, has $Q \sim 10^5$. Advanced nanofabricated mechanical oscillators can reach $Q \sim 10^9$. The best mechanical oscillators we know of are astrophysical, and are stable by virtue of their size. For example, the rotation of the Earth around the Sun has $Q \sim 10^7$, with the damping due to interaction with the Moon. The rotation of a neutron star has $Q \sim 10^{10}$, damped by the emission of gravitational waves.

The high Q factors of atomic systems have led to the development of extremely precise atomic clocks. Another key advantage of atomic systems is, of course, that all atoms are identical, allowing the clocks to be reproduced anywhere.

1.3 Lasers

In this section, we briefly describe lasers.

- A laser consists of a gain medium and two end mirrors, called the output coupler and high reflector, with reflectivities R_1 and R_2. Define the gain coefficient $\gamma(\omega)$ of the gain medium by

$$\frac{dI}{dz} = \gamma(\omega)I(z).$$

If the medium is at thermal equilibrium at any positive temperature, then absorption beats stimulated emission, giving a negative gain coefficient, and hence no lasing.

- Lasing is only possible if some higher energy state has a greater probability of occupancy than a lower energy state, a situation known as population inversion. In that case, one can show

$$\gamma(\omega) = \frac{\lambda^2}{4n^2\tau}\Delta N\, g(\omega), \quad \Delta N = N_2 - \frac{g_2}{g_1}N_1$$

  where n is the index of refraction and τ is the lifetime of the upper level.

- A typical laser design is a four-level laser, where the levels decay as

$$3 \xrightarrow{\text{fast}} 2 \xrightarrow{\text{slow}} 1 \xrightarrow{\text{fast}} 0.$$

  A pump drives $0 \to 3$, and population inversion occurs in the $2 \to 1$ transition. The gain coefficient increases with the pumping power, with lasing beginning at the laser threshold.

- A simpler option is a three-level laser, where the states 0 and 1 coincide. However, in this case the laser threshold is higher, because one has to significantly reduce the occupancy of the 0 state to achieve population inversion.

- In the steady state, we have

$$R_1 R_2 \xi e^{2\gamma L} = 1$$

  where ξ accounts for all other losses, and L is the length of the cavity.

- The laser beam is extremely narrow because it is generated by light that has bounced repeatedly between the cavity mirrors. The transverse mode structure of the beam is given by

$$E_{mn}(x,y) \propto H_m(\sqrt{2}x/w)H_n(\sqrt{2}y/w)e^{-r^2/w^2}$$

  where the H_n are the Hermite polynomials and w is the beam spot size, which determines the width of the beam. Usually one tries to operate the laser in the $n = m = 0$ mode, since this mode has the minimum divergence, with $E \propto e^{-r^2/w^2}$. The modes that propagate out of the cavity depend especially on the size and shape of the aperture.

- There is also nontrivial longitudinal mode structure. Since the longitudinal mode is a standing wave, the frequency spacing is

$$\Delta\omega = \frac{\pi c}{nL}.$$

  Normally, the mode spacing is much smaller than the linewidth, so multiple longitudinal modes can be excited at once if the broadening is inhomogeneous; the excitations of these modes have randomly related phases. This is called multi-mode operation. However, it means that the effective width of the laser is just as wide as the spectral line itself.

- To achieve the longest possible coherence lengths, one needs single-mode operation. This can be done by placing a frequency selective element such as a Fabry–Perot etalon into the cavity, damping all other modes. By tuning the etalon, one can scan which mode to excite.

- Finally, one can operate the laser as mode-locked. In this case, multiple longitudinal modes are excited, but their phases are locked together. The regular frequency spacing of the modes causes a regular train of pulses separated in time by $2nL/c$, while the length of each pulse is limited by the inverse linewidth. The linewidth can hence be used to generate short pulses. Physically, one can think of mode-locked operation as having just a single pulse bouncing back and forth in the cavity, with an output pulse emitted every time it hits the output coupler.

- Lasers are characterized by the gain medium used, the design of cavity, and the mode of operation. The gain medium determines the frequencies generated, while the cavity design influences the transverse and longitudinal structure. The gain medium can be a gas, a solid, or even in rare cases a liquid. Conventional lasers have output wavelengths ranging from the UV to the far IR; more exotic techniques are used to make lasers far outside this range.

2 Quantum Optics

2.1 Background

In this section we cover quantum optics, meaning phenomena that must be modeled by the interaction of atoms with quantized fields. First, we cover the history of quantum optics.

- A timeline of developments in quantum optics is given below.

Year	Authors	Development
1901	Planck	Theory of black-body radiation
1905	Einstein	Explanation of the photoelectric effect
1909	Taylor	Interference of single quanta
1909	Einstein	Radiation fluctuations
1927	Dirac	Quantum theory of radiation
1956	Hanbury Brown and Twiss	Intensity interferometer
1963	Glauber	Quantum states of light
1972	Gibbs	Optical Rabi oscillations
1977	Kimble, Dagenais, and Mandel	Photon antibunching
1981	Aspect, Grangier, and Roger	Violations of Bell's inequality
1985	Slusher *et al.*	Squeezed light
1987	Hong, Ou, and Mandel	Single-photon interference experiments
1992	Bennett, Brassard *et al.*	Experimental quantum cryptography
1995	Turchette, Kimble *et al.*	Quantum phase gate
1995	Anderson, Wieman, Cornell *et al.*	Bose–Einstein condensation of atoms
1997	Mewes, Ketterle *et al.*	Atom laser
1997	Bouwmeester *et al.*, Boschi *et al.*	Quantum teleportation of photons
2002	Yuan *et al.*	Single-photon light-emitting diode

- In the early 1900s, Planck and Einstein developed the idea of photons via blackbody radiation. In 1909, Taylor performed the double slit experiment with low enough intensities so that only one photon at a time would be in the apparatus. These experiments were taken as evidence for photons, but in reality, they only require that *something* is quantized; as mentioned above, they can be explained using quantized matter and classical light.

- In the 1910s, the standard formulation of quantum mechanics was developed, culminating in Dirac's quantization of the electromagnetic field in 1927.

- During the following decades, the full theory of quantum electrodynamics was developed, but only a handful of physical effects at the atomic scale really went beyond the semiclassical approach. Two examples were spontaneous emission and Lamb shift, phenomena which are tied to the quantum vacuum.

- In 1956, Hanbury-Brown and Twiss (HBT) demonstrated a counterintuitive effect based on correlations between starlight intensities recorded on two separated detectors, which could be explained using the quantized electromagnetic field. While it was later found that this effect could be explained semiclassically, this was one of the founding experiments of quantum optics.

- The laser was invented in 1960, and it was expected that the resulting light would differ from classical light from incoherent sources. Glauber formulated the theory of optical coherence in 1963, and demonstrated that certain states of light could have different statistical properties than that of classical light. In 1977, the resulting prediction of photon antibunching was observed in the lab. In 1985, squeezed states were produced in the lab.

- In the following few decades, the field of quantum optics rapidly expanded, providing insight into the interpretation of quantum mechanics, and tools for quantum computation, quantum information, and atomic physics.

Note. The fact that the photoelectric effect and its relatives do not provide direct evidence for photons may be somewhat of a shock, since almost all non-optics textbooks say otherwise. The reason for this difference in nomenclature is a difference in culture.

Particle physicists, for instance, are aware that a semiclassical theory with some quantized objects and some classical ones is ultimately inconsistent. For example, the early BKS theory of semiclassical light notoriously required violations of energy conservation. Depending on the setup, one is forced to lose conservation laws, or violate the uncertainty principle, or even lose causality. Problems like these are why the particle physics community has been searching for a quantum theory of gravity for decades, even though semiclassical gravity suffices to explain all observations to date, and for the foreseeable future. In the end, everything must be quantized, so semiclassical theory is not a good place to stop.

On the other hand, optical physicists have a strong trust in Maxwell's equations, from nearly two centuries of experience. Many regard quantum field theory as an extreme mathematical complication that should not be invoked without good reason – anything that can be explained with Maxwell's equations alone should be. This difference in culture often leads to bitter disputes, e.g. see Lamb's famous Anti-photon paper or this discussion. In these notes, we will take the latter approach, striving to understand as much as possible semiclassically, but we will ultimately be led by necessity to a fully quantum theory of light.

Next we establish some basic facts about coherence.

- At the simplest level, a light beam can be monochromatic and hence perfectly coherent, or "completely incoherent", displaying no interference effects. These are two common limits, but many real sources are partially coherent, so we would like a quantitative measure of coherence.

- For example, sunlight was thought to be completely incoherent in the early 1800s, but it was demonstrated that it could produce fringes in a double slit experiment if the spacing was less than about 0.05 mm. More generally, in a real experiment we will never have perfect interference; fringes simply have more or less contrast depending on the coherence properties of the light.

- We can consider the electric field as a function of time at a given point. For any finite light wave, the Fourier transform will have a width $\Delta\omega$. There is a corresponding coherence time $\tau_c \sim 1/\Delta\omega$. In practice, what this means is that knowledge of the phase at some time only allows one to predict the phase accurately within a coherence time later.

- For example, one very crude model of finite coherence time is that the source is a perfect sinusoid within a coherence time, but every time τ_c its phase randomly changes.

- This notion is called temporal coherence, but there is an independent notion of spatial coherence. The spatial coherence length is the transverse separation within which the phases at two points will be correlated. (Similarly, one can think of temporal coherence as "longitudinal coherence", with an associated coherence length $c\tau_c$.)
- For example, in practice this means that fringes in a double slit experiment will be visible if the slits are within the spatial coherence length. On the other hand, suppose we run the double slit experiment with the slits well within the spatial coherence length. We can effectively delay the light from one of the slits by placing a piece of glass in front of it. Then the fringes will be visible if the change in optical path length is within the temporal coherence length.
- Consider a typical example of thermal emission of light, such as from an incandescent lamp, where many atoms emit light independently. It would seem that the light is completely incoherent. However, if we filter the light with a narrow filter, the temporal coherence increases.
- Furthermore, the spatial coherence length increases the further one gets from the lamp. This is because the fields at two nearby, transversely separated points are composed of the fields of all the emitters times appropriate path length phase factors, and for large separation L, the path length differences fall as $1/L$ by the standard far-field formulas. Another way to think about this is that initially jagged wavefronts smooth out as the distance increases. In optics, this point can be formalized using the van Cittert–Zernike theorem.
- Sunlight has a very large bandwidth, and hence a tiny temporal coherence length. But as we stated above, it has a measurable spatial coherence length, and the previous point is the reason why. (Lasers can also have sizable coherence times, on the order of seconds; this is extremely long on atomic timescales!)

Coherence can be defined more quantitatively in terms of correlation functions. In the following, we ignore polarization effects, treating the light wave as a scalar E.

- For simplicity, we'll focus on temporal coherence. This can be quantified using the first-order correlation function
$$g^{(1)}(\tau) = \frac{\langle E^*(t)E(t+\tau)\rangle}{\langle |E(t)|^2\rangle}$$
where the averages are time averages, which by ergodicity are assumed to match ensemble averages. That is, $g^{(1)}(\tau)$ is simply the normalized autocorrelation function.
- The quantity $g^{(1)}(\tau)$ is called the first-order correlation function because it depends directly on the first power of the electric field. Perfectly coherent light has $|g^{(1)}(\tau)| = 1$ for all τ, while partially coherent light has a falloff after a coherence time τ_c.
- There are many more functions one can define. For example, the analogous quantity in space,
$$g^{(1)}(x_1, x_2) = \frac{\langle E^*(x_1)E(x_2)\rangle}{\sqrt{|E(x_1)|^2|E(x_2)|^2}}$$
is called the mutual coherence function.
- In our crude model of partial coherence as essentially monochromatic light, but with the phase randomly changing every time τ_c, the magnitude of the first-order correlation function decreases linearly from one to zero over time τ_c, then remains at zero.

- This calculation can be refined. The Wiener–Khinchin theorem states that the Fourier transform of the first-order correlation function is the power spectral density of $E(t)$. Thus we can compute it properly using known lineshapes.
- For example, for light with a Lorentzian lineshape of half width $\Delta\omega$,

$$g^{(1)}(\tau) = e^{-i\omega_0\tau}e^{-|\tau|/\tau_c}, \quad \tau_c = \frac{1}{\Delta\omega}.$$

  For a Gaussian lineshape with standard deviation $\Delta\omega$,

$$g^{(1)}(\tau) = e^{-i\omega_0\tau}\exp\left(-\frac{\pi}{2}\left(\frac{\tau}{\tau_c}\right)^2\right), \quad \tau_c = \frac{\sqrt{8\pi\log 2}}{\Delta\omega}.$$

- In interferometry, the visibility of a fringe is

$$V = \frac{I_{\max} - I_{\min}}{I_{\max} + I_{\min}}.$$

  Suppose we are using a Michelson interferometer, which is based on time delays τ. The visibility hence depends on the temporal coherence. The output field is

$$E^{\text{out}}(t) = \frac{1}{\sqrt{2}}(E(t) - E(t+\tau))$$

  and straightforward computations give

$$I(\tau) \propto \langle E^{\text{out}*}(t)E^{\text{out}}(t)\rangle \propto 1 - \text{Re}(g^{(1)}(\tau)).$$

- For reasonably coherent light, the phase of $g^{(1)}(\tau)$ changes rapidly compared to its magnitude, so we can treat its magnitude as constant over one interference fringe. For that fringe, we have

$$I_{\max} = 1 + |g^{(1)}(\tau)|, \quad I_{\min} = 1 - |g^{(1}(\tau)|, \quad V = |g^{(1)}(\tau)|.$$

Next, we cover the very basics of nonlinear optics.

- In the optical regime, $\mu_r \approx 1$ since the relevant magnetic moments do not react fast enough to keep up with the fields. **(understand this timescale better)** On the other hand, ϵ_r can still be sizable because the fastest polarization response comes from electronic configurations, whose characteristic frequency is optical.
- In a linear medium, the polarization is proportional to the applied field. More generally,

$$P = \epsilon_0(\chi^{(1)}E + \chi^{(2)}E^2 + \chi^{(3)}E^3 + \ldots) = P^{(1)} + P^{(2)} + P^{(3)} + \ldots.$$

  The $\chi^{(n)}$ for $n > 1$ are called the n^{th} order nonlinear susceptibilities, and the $P^{(n)}$ are called the n^{th} order nonlinear polarizations. These higher-order terms are studied in nonlinear optics.
- As an example, if a medium is excited at frequencies ω_1 and ω_2, then the second-order nonlinear polarization has frequencies $\omega_1 \pm \omega_2$, as well as $2\omega_1$ and $2\omega_2$. Second-order effects also allow a photon at a "pump" frequency ω_p to be downconverted into two photons, at the signal and idler frequencies ω_s and ω_i, where $\omega_s + \omega_i = \omega_p$.

- Yet another example is parametric amplification, which has close links to parametric resonance. In this case, the medium experiences a weak signal field at ω_s and a powerful pump field at ω_p. Idler photons are produced at frequency $\omega_p - \omega_s = \omega_i$. These idler photons interact with the pump field again to produce photons at frequency $\omega_p - \omega_i = \omega_s$, amplifying the signal. There is also the degenerate case $\omega_s = \omega_i = \omega_p/2$ where signal photons are made directly. This is a phase-sensitive amplifier, as whether it amplifies or deamplifies the signal depends on the signal and pump phases.

- More generally, the nonlinear susceptibilities should be described by tensors. For instance,

$$P_i^{(2)} = \epsilon_0 \chi_{ijk}^{(2)} E_j E_k.$$

  Note that $\chi_{ijk}^{(2)}$ must be symmetric in its last two indices. Also note that the second-order susceptibility of a material with parity/inversion symmetry is automatically zero. In particular, all isotropic media, such as gases, liquids, and glasses, have inversion symmetry due to their lack of long range order. In these cases, the leading effect is the third-order susceptibility.

2.2 Photon Statistics

In quantum optics, an apparently continuous beam of light is actually comprised of discrete photons. This means that photon detection statistics are one way to probe quantum optical effects.

- A photon counter consists of a very sensitive light detector such as a photomultiplier tube (PMT) or avalanche photodiode (APD). The detector produces short voltage pulses in response to the light beam, which are counted.

- Since these counts are discrete, we expect to see fluctuations in their number over time, just like clicks in a Geiger counter. The tricky part is that these fluctuations can come from both the photon statistics of the light beam, and the statistical nature of the photodetection process, since the detector is quantum as well. We must find a way to distinguish these two.

- The photon flux Φ is defined as the average rate of photons passing through a cross-section of the beam, and

$$\Phi = \frac{IA}{\hbar\omega} = \frac{P}{\hbar\omega}.$$

  Photon-counting detectors are specified by their quantum efficiency η, which is the ratio of the number of photocounts to the number of incident photons. Hence the average count rate is $\mathcal{R} = \eta\Phi$. Modern detectors typically have $\eta \sim 10\%$, and up to $\eta \sim 50\%$ in certain wavelengths.

- Photon-counting detectors also have a "dead time" of order $1\,\mu$s that must elapse between successive counts, placing a bound on the $\mathcal{R}$ that can be reliably measured. In practice, one needs very faint light beams, with optical powers of $10^{-12}\,$W or less.

- In order to focus on the statistics of the light beam, we ignore any issues with photodetection for now. We simply assume that the number of counts reported in some time is simply the result of applying a photon number measurement to some subset of the beam.

- First we consider perfectly coherent monochromatic light, which therefore must have constant intensity. We know that within a beam segment of length L,

$$\overline{n} = \frac{\Phi L}{c}.$$

  Now consider splitting the beam into tiny segments. Within each segment, there are either zero or one photons. Furthermore, as we will show properly later, the counts in each segment are independent. These two statements are enough to fix the probability distribution to be Poisson,

$$P(n) = \frac{\overline{n}^n}{n!} e^{-\overline{n}}.$$

  The $\overline{n}^n$ accounts for the probability of having photons in n segments, while the $n!$ in the denominator prevents overcounting.

- Let Δn be the standard deviation. Then for Poissonian statistics,

$$\Delta n = \sqrt{\overline{n}}.$$

  We say that photon statistics are sub-Poissonian if $\Delta n < \sqrt{\overline{n}}$, and super-Poissonian if $\Delta n > \sqrt{\overline{n}}$.

- It is not difficult to imagine sources of light that are super-Poissonian. For example, any classical fluctuations in the intensity would increase Δn. On the other hand, it is difficult to imagine a source of light that is sub-Poissonian, because it would have a narrow distribution than perfectly coherent light, which is the most stable possible classical light source.

- Light with sub-Poissonian statistics exists, and is thus called non-classical light. One extreme example would be a photon number state, where $\Delta n = 0$.

Example. Thermal light. Since thermal light has a fluctuating intensity, we expect it has super-Poissonian statistics. To check this, consider the distribution of photons in a given mode of light in thermal equilibrium. Then

$$P(n) = \frac{e^{-n\hbar\omega/k_B T}}{Z}.$$

Using some standard geometric series tricks, and defining $x = e^{-\hbar\omega/k_B T}$, we have

$$\overline{n} = \frac{x}{1-x} = \frac{1}{e^{\hbar\omega/k_B T} - 1}$$

which is of course the Bose–Einstein distribution. We can use this to write the probability distribution in a nicer form,

$$P(n) = \frac{1}{\overline{n}+1} \left(\frac{\overline{n}}{\overline{n}+1} \right)^n.$$

By another application of geometric series tricks,

$$(\Delta n)^2 = \overline{n} + \overline{n}^2.$$

Einstein had derived this by 1909, using solely Planck's law and the general result

$$\operatorname{var} E = k_B T^2 C_V$$

which we derived in the notes on Undergraduate Physics. The first term is due to quantum effects, while the second is due to classical intensity fluctuations, and called wave noise. The wave noise will eventually disappear with enough averaging. For example, if we couple to N distinct thermal modes with similar occupancies, then the factor of $\overline{n}^2$ above becomes $\overline{n}^2/N$. In the limit of coupling to many modes or waiting for a long time, the statistics become approximately Poissonian.

Example. Chaotic light. Consider classically partially coherent light, with intensity fluctuations on a timescale τ_c. Then in a time interval T,

$$(\Delta n)^2 = \langle W(T) \rangle + \langle \Delta W(T)^2 \rangle$$

where $W(T)$ is the count rate for an interval T,

$$W(T) = \int_t^{t+T} \eta \Phi(t')\, dt'.$$

For $T \lesssim \tau_c$ the standard deviation is hence significantly increased, as we'd expect.

Now we account for the physics of the photodetector.

- Above, we have worked entirely with photons, but called some of the photon states "classical". This is because their effects can be mimicked in the semiclassical theory by other effects in the detection process.
- First, it is worth noting that losses can degrade sub-Poissonian photon statistics, even purely classically. Suppose we model a loss as an independent chance to lose each photon. This adds an additional source of fluctuations, by random sampling. In the limit where these effects are strong, even a number eigenstate can be turned into a Poisson distribution, since each photon has an independent small chance of getting through.
- Physically, this loss can be due to inefficient collection optics, where only a fraction of the light emitted from the source is collected, losses in the optical components, and inefficiency in the detection process.
- Worse, quantum effects in the detector also produce Poissonian statistics, even with classical light. Let us model the light as classical with constant intensity. By an application of Fermi's golden rule, we expect that the probability of emission of a photoelectron is constant per small time interval, and that photoemission events are statistically independent of each other. These two rules automatically mean that the number of emissions in a given time period is Poissonian.
- For classical light with time-varying intensity, the counts become super-Poissonian, by the same logic as above. However, they cannot be sub-Poissonian, so an observation of sub-Poissonian statistics *requires* quantum optics to explain. Of course, a full treatment using quantum optics requires quantizing both the light and the detector.
- If one does this, one can show that the photocount number N has variance

  $$(\Delta N)^2 = \eta^2 (\Delta n)^2 + \eta(1-\eta)\overline{n}$$

  where the quantum efficiency is $\eta = \overline{N}/\overline{n}$. If $\eta = 1$, then we have $\Delta N = \delta n$ as expected. But if either η is low, or n is Poissonian, then N is Poissonian. Therefore, it is a pressing issue to find detectors with high η.

Now we discuss a different way of detecting photons, which works at high intensity and typically allows higher η, up to 90%. The principles are the same, but the language is slightly different.

- The detectors above multiplied the input, turning individual photons into current pulses; the fluctuations then came from the statistics of the pulses. An alternative strategy is to use a photodiode, where each individual photon creates one photoelectron, which is not subsequently multiplied. For any non-weak beam, the resulting current is high enough that one sees an average photocurrent
$$\langle i \rangle = \eta e \Phi$$
and the quantum nature of the system manifests in small fluctuations Δi of this current.

- Working in the semiclassical picture, suppose the light is perfectly coherent. Then the resulting electron statistics are expected to be Poissonian, $(\Delta N)^2 = \langle N \rangle$, which means that $(\Delta i)^2 \propto \langle i \rangle$.

- Furthermore, the current noise has zero autocorrelation function, so by the Wiener–Khinchin theorem it is white noise, with a constant power spectral density proportional to $\langle i \rangle$. This noise is called quantum noise or shot noise, and the resulting limit on signal detection is called the standard quantum limit.

- We can measure this noise by attaching the output current of the photodiode to a high-pass filter (to remove the constant component $\langle i \rangle$), then to an amplifier. In practice, the noise power spectral density will be high at low frequencies due to classical fluctuations, such as vibrations of the laser cavity. At intermediate frequencies, the spectrum is flat due to shot noise. At high frequencies, the spectrum falls off to zero simply because of the finite response time of the photodiode.

- The low-frequency classical noise can be removed in many ways, depending on its source. For example, if the input has a classically time-varying intensity, then a "noise eater" setup with an active negative feedback system can smooth it out. Another idea is the "balanced detector", where one splits a beam in half, uses one as a "control" and passes the other through a sample, and subtracts the intensities of the two at the end. Neither of these can do anything about the shot noise, which is intrinsic to the photon statistics.

- In the semiclassical picture, it is impossible to beat the standard quantum limit, because it occurs inevitably from the interaction of classical light with a quantized detector. Hence shot noise has been a difficult issue in many practical applications. However, the standard quantum limit can be beaten using non-classical states of light. Such developments are barely 15 years old, and have already been put to use in precision physics experiments, such as LIGO.

We now discuss how sub-Poissonian photon statistics can be observed in the lab.

- One ingenious method is to drive a high-efficiency light emitter with a current source with sub-Poissonian electron statistics. For example, a Franck–Hertz tube quickly creates one photon per electron, at a timescale faster than fluctuations in the current intensity. Then the resulting photon statistics simply track the electron statistics.

- In order to create sub-Poissonian electron statistics, one can "boil" electrons off a hot cathode. For individual electrons, this would yield Poissonian statistics. However, over time a space charge develops around the cathode, causing the electron flow to be more regular.

- The obvious reason that it is easier to make sub-Poissonian electron statistics is that electrons repel each other. But a deeper reason is that electrons are fermions while photons are bosons, and bosons like to clump up.
- Experiments along these lines were performed in the 1990s. With such a setup, there are many points of inefficiency, leading to only very slightly sub-Poissonian statistics, but enough to serve as a proof of principle.
- Later experiments used solid-state emitters such as LEDs or laser diodes (LDs), which have higher efficiencies. This again works because they produce photons from electrons that have sub-Poissonian statistics.
- To quantify the statistics, it is convenient to use the Fano factor

$$F = \frac{\text{measured noise}}{\text{shot noise limit}}.$$

If the entire system converts electrons from the source into signal with efficiency η, then the measured Fano factor is

$$F = \eta F_d + (1 - \eta)$$

where F_d is the Fano factor of the electron source.

2.3 Correlation Functions

We discussed photon statistics phenomenologically above, from the standpoint of what one directly sees when counting photons in the lab. We now give a more abstract, but more flexible account using correlation functions. We begin with a historically important experiment that is naturally explained in this formalism.

- The Michelson interferometer was developed to measure the angular diameters of distant astronomical bodies, by effectively measuring their spatial coherence. If their diameters can be estimated otherwise, one hence has a measurement of the distance to the star.
- The basic idea is to use stellar light as the source for a double slit experiment. For simplicity, we suppose this light is close to monochromatic. First consider a distant point source of light. It can be made to hit two slits separated by distance a, and hence creates a double slit pattern.
- Now consider two point sources of light, close to each other, but completely incoherent with respect to each other. Then the intensity pattern on the screen is just the sum of their individual intensity patterns. If the slits are "pointed" towards one of the sources, then the other arrives off-axis, and hence with a phase shift, causing a shift in the resulting fringe location.
- In particular, for small a, one still has a perfectly good double slit interference pattern, even though we are using light from two sources that are totally incoherent. The situation only changes as we change a, as in this case the maxima of one can overlap the minima of the other.
- More generally, for a distant star of finite size, we must sum over all of the intensity patterns produced by each point on the surface of the star. This ends up analogous to diffraction from a circular aperture, as in both cases we sum up phasors with phases determined by their heights on a disc. The fringes disappear for the first time when $a = (1.22)\lambda/\theta$ where θ is the angular diameter of the star, where the 1.22 is again a zero of a Bessel function.

- As a general principle, diffraction effects limit the angular resolution of telescopes to order λ/a where a is their radius. The point of the Michelson interferometer is that one can have a large a without building a prohibitively large telescope, by simply using two widely separated telescopes. The light from the two is fed to the double slit apparatus using systems of mirrors. (Of course one still needs a big telescope to get enough intensity; that is a separate problem.)
- However, a is still required to be very large in order to resolve all but the largest and nearest stars. For example, Michelson originally used his apparatus to measure the size of Betelgeuse, thereby discovering red giants. However, as a increases, it becomes more difficult to make a sufficiently stable mirror system.
- In 1956, Hanbury-Brown and Twiss came up with a refinement that avoided this issue. We note that the reason that light entering the two slits makes an interference pattern at all is because the fields at each slit are correlated. The signal is significantly reduced for $a \gtrsim \lambda/\theta$ because this is the spatial coherence length of the light from the star.
- Therefore, there is no reason we actually have to interfere the light in real time! In principle, one can simply record the field amplitudes for the telescopes separately, then compute the appropriate correlation functions between them later.
- Of course, the reason that interferometry is useful at all is that in practice, the field amplitudes vary too fast to be tracked directly; interferometry transforms this into a relatively static intensity pattern. However, it turns out that recording the *intensities* at each telescope and correlating those gives enough information. This technique is called correlation interferometry.
- It is intuitive why correlating the intensities should provide some information. For small a, we have $I_1(t) = I_2(t)$, so their correlation is maximal. For larger a, $I_1(t) \neq I_2(t)$ but the two individually retain the same statistics, so the correlation function falls.
- Today, interferometric techniques are ubiquitous in astronomy. For example, the Event Horizon Telescope used correlations between telescopes with separation on the order of the Earth's radius to image a black hole; this technique is known as very-long baseline interferometry. One of the technical challenges is maintaining timing consistency between the telescopes, which is achieved with atomic clocks.
- The original HBT effect was controversial, because the two "telescopes" it used were actually single photon counters, and HBT studied correlations between photon counts. However, for the expected correlations to exist, photon arrival at the two detectors would have to be correlated, contradicting the naive picture of photon emission as a completely random process.
- Some clear thought shows that these objections are invalid. For example, if it really were the case the photon arrival at the two detectors could not be correlated, then classical intensities could never be correlated either, since these merely emerge in the limit of many photons. Furthermore, photon emission from the star is not completely random, in the intuitive classical sense, because the amplitude to emit a photon is increased by the number of photons already present, a typical bosonic effect. That is, the photons coming from the star are "bunched" together.
- The HBT experiment can also be understood semiclassically, thinking about how the quantum detectors respond to classical fluctuations in intensity. This is simply because HBT used detectors with low quantum efficiencies, and thermal sources. However, the HBT effect is a

cornerstone of quantum optics because, if one improves the detectors and uses nonthermal sources, one can demonstrate effects that are clearly impossible in the semiclassical framework.

- To demonstrate their point, HBT repeated their experiment by correlating the intensity fluctuations of nearly monochromatic light split by a half-silvered mirror, with one path receiving an extra delay. That is, they computed $\langle \Delta I_1(t) \Delta I_2(t+\tau) \rangle$. Using such a setup allows one to probe temporal coherence, which is somewhat simpler than spatial coherence. For that reason, we'll focus on temporal coherence below.

We now quantify the above with the second-order correlation function, initially semiclassically.

- We define the second-order correlation function by

$$g^{(2)}(\tau) = \frac{\langle E^*(t) E^*(t+\tau) E(t+\tau) E(t) \rangle}{\langle |E(t)|^2 \rangle \langle |E(t+\tau)|^2 \rangle} = \frac{\langle I(t) I(t+\tau) \rangle}{\langle I(t) \rangle \langle I(t+\tau) \rangle}.$$

  For simplicity we assume the source has constant average intensity,

$$\langle I(t) \rangle = \langle I(t+\tau) \rangle \equiv \langle I \rangle.$$

- Defining $I(t) = \langle I \rangle + \Delta I(t)$, we have

$$\langle I(t) I(t+\tau) \rangle = \langle I \rangle^2 + \langle \Delta I(t) \Delta I(t+\tau) \rangle$$

  which implies that

$$g^{(2)}(\tau) = 1 + \frac{\langle \Delta I(t) \Delta I(t+\tau) \rangle}{\langle I \rangle^2}.$$

  For *any* possible time-dependence of $I(t)$, we have

$$g^{(2)}(0) \geq 1, \quad g^{(2)}(0) \geq g^{(2)}(\tau)$$

  where the latter follows from the Cauchy–Schwarz inequality.

- For example, for perfectly coherent, monochromatic light the intensity is constant, so $g^{(2)}(\tau) = 1$. Also, it can be shown that if the spectral line is Doppler-broadened, and hence Gaussian,

$$g^{(2)}(\tau) = 1 + \exp\left(-\pi(\tau/\tau_c)^2\right)$$

  where τ_c is the coherence time. Moreover, a Lorentzian lifetime-broadened source has

$$g^{(2)}(\tau) = 1 + \exp\left(-2|\tau|/\tau_0\right)$$

  where τ_0 is the lifetime or collision time. Note that in both of the latter cases, we have $g^{(2)}(0) = 2$.

- Since HBT detected individual photons, they measured the quantum mechanical analogue,

$$g^{(2)}(\tau) = \frac{\langle n_1(\tau) n_2(t+\tau) \rangle}{\langle n_1(t) \rangle \langle n_2(t+\tau) \rangle}.$$

  Their measurements confirmed that it behaved as expected semiclassically, with $g^{(2)}(0) = 2$. This specific result is sometimes called the HBT effect.

- However, now that we are thinking of light quantum mechanically, it is clear that we can get results that clearly have no classical explanation. For example, suppose we send in one photon at a time through the apparatus. Then the number counts at the detectors are perfectly anticorrelated, $g^{(2)}(0) = 0$!
- This violates our general semiclassical result above, and also our semiclassical intuition. While detectors with $\tau = 0$ can easily be correlated, by responding to correlated classical fluctuations, they can't be anticorrelated, as this would require some kind of nonlocal conspiracy.
- We describe light as

  $$\text{bunched/chaotic: } g^{(2)}(0) > 1, \quad \text{random/coherent: } g^{(2)}(0) = 1, \quad \text{antibunched: } g^{(2)}(0) < 1.$$

  Heuristically, the photons are spaced randomly in coherent light, while they are spaced more regularly in antibunched light. Note that this is closely related to photon counting statistics. Randomly spaced photons correspond to Poissonian light, while antibunched light can be sub-Poissonian. The main difference is that observing antibunching is more saliently quantum than observing sub-Poissonian statistics.
- At a high level, the reason that bunched light is easier to produce than antibunched light is that photons are bosons, while like to clump up. One could say that it's because classical light is bunched, but the only reason that we can think of light as a classical field in the first place is again that photons are bosons.
- Note that antibunched light is not necessarily sub-Poissonian. For example, $g^{(2)}(\tau)$ could quickly rise above 1, then eventually decay back to 1. However, in practice the two properties usually occur together.
- As we saw above, the archetypical antibunched light comes from single photon sources. The first experimental demonstration of antibunched light was performed by Kimble et al. in 1977. They took advantage of the fluorescence of atomic sodium. By shining a weak laser beam on a sample, a small number of sodium atoms were excited, which subsequently took a long time to decay from a metastable state. This yielded widely spaced single photons. Today, antibunching has been observed from a wide variety of emitters.
- Antibunching experiments are connected to the development of single-photon sources of light. These are useful for defending quantum communication protocols against photon number splitting attacks. Single-photon sources using single quantum dots are currently under development. One can also generate single photons by strongly diluting a coherent state, but this isn't really the same thing, since we either have a great chance of getting no photons at all, or a sizable chance of getting more than one.

Example. Suppose one has a source that emits pulses of n photons. If these pulses are passed through a 50/50 beam splitter as in the HBT experiment, then the probability of k photons in the first arm is

$$p_k = \frac{1}{2^n}\binom{n}{k}.$$

This is classically intuitive, but we will justify it carefully below. As a result, we have

$$g^{(2)}(0) = \frac{2^{-n}}{(n/2)^2}\sum_{k=0}^{n}\binom{n}{k}n(n-k).$$

This can be simplified using the identities

$$\sum_{k=0}^{n} \binom{n}{k} k = 2^{n-1} n, \quad \sum_{k=0}^{n} \binom{n}{k} k^2 = 2^{n-2} n(n+1)$$

which yields

$$g^{(2)}(0) = 1 - \frac{1}{n}.$$

The result is still antibunched as expected, though less so, because the law of large numbers starts to kick in.

2.4 Quantum Modes

We now put what we have discussed on firmer ground by quantizing the electromagnetic field.

- Contrary to what we might expect from the notes on Quantum Field Theory, this is not a difficult task, because we don't care about formal issues like Lorentz invariance and gauge invariance. In fact, for almost all the applications we will discuss, it suffices to focus on a single mode of the field, which is simply a 1D harmonic oscillator. We won't even care much about how this mode is distributed in space.

- The electric field amplitude for a mode is

  $$\hat{E} = \mathcal{E}_0(\hat{a} + \hat{a}^\dagger) f(x)$$

  where $f(x)$ is a field profile, $\mathcal{E}_0$ is a normalization constant which ensures that the field energy of a one-photon state is $\hbar\omega$, and we have ignored the vector nature of the field entirely. From this point on we suppress the hats.

- We see that the electric field amplitude can be thought of as the "position" of this mode oscillation. Since the magnetic field is 90° out of phase, it can be thought of as the "momentum".

- Instead of working with (E, B) directly, it is convenient to use the quadrature operators to remove unwanted normalization constants,

  $$X_1 = \frac{1}{2}(a + a^\dagger), \quad X_2 = \frac{1}{2i}(a - a^\dagger), \quad [X_1, X_2] = \frac{i}{2}.$$

  These are related to the familiar position and momentum operators by

  $$X_1 = \sqrt{\frac{\omega}{2\hbar}}\, q, \quad X_2 = \frac{1}{\sqrt{2\hbar\omega}} p.$$

  The uncertainty principle for quadrature operators is

  $$(\Delta X_1)(\Delta X_2) \geq \frac{1}{4}.$$

- It will be useful to visualize a quantum state as a probability distribution for an ensemble on the (X_1, X_2) phase space; we will make this more technically precise below. For example, a coherent state $|\alpha\rangle$ can be roughly described as a circle in this phase space centered at α with diameter $1/2$. A thermal state is a circle centered at the origin. Intuitively, we expect a number state $|n\rangle$ to look like a ring in phase space, with definite radius n but uniform phase.

- It is useful to think of phase space in polar coordinates. Then the phase corresponds to the phase ϕ of the wave, and the length corresponds to the square root of the number of photons.

Note. It is infamously tricky to define the "phase operator" $\hat{\phi}$. The most obvious issue is that ϕ is only defined up to 2π, so it's better to try to work with $e^{i\phi}$ as a whole, then use that to find the desired results. The first attempt was made by Dirac, who defined

$$a = e^{i\phi}\sqrt{n}, \quad a^\dagger = \sqrt{n}e^{-i\phi}$$

where ϕ is assumed to be a Hermitian operator. Then $[a, a^\dagger] = 1$ implies

$$e^{i\phi}ne^{-i\phi} - n = 1$$

which corresponds to

$$[e^{i\phi}, n] = e^{i\phi}.$$

Expanding the exponential to first order gives the uncertainty relation,

$$[n, \phi] = i, \quad \Delta n\, \Delta\phi \geq \frac{1}{2}.$$

This all looks quite promising, but it doesn't work. The easiest route is to note that if ϕ is Hermitian, then $e^{i\phi}$ is unitary. In the number basis,

$$e^{i\phi} = \begin{pmatrix} 0 & 1 & 0 & \dots \\ 0 & 0 & 1 & \dots \\ 0 & 0 & 0 & \dots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}, \quad e^{-i\phi} = \begin{pmatrix} 0 & 0 & 0 & \dots \\ 1 & 0 & 0 & \dots \\ 0 & 1 & 0 & \dots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}.$$

Just like how the exponential of p translates in x, the exponential of ϕ translates in number. But then we have $e^{i\phi}e^{-i\phi} \neq 1$, because the left-hand side annihilates the vacuum. Fundamentally, the problem is that the number operator has a "bottom rung".

Barnett and Pegg have proposed to define unphysical "negative number states", and let

$$e^{i\phi} = \sum_{n\in\mathbb{Z}} |n\rangle\langle n+1|.$$

Other authors have proposed to simply accept the nonunitarity of $e^{i\phi}$, since it is "small" as long as we work at high occupancies. No matter what route we choose, any phase operator ϕ still has the problem of being defined only up to 2π, which leads to pathologies. One possible resolution is to simply work with

$$\Phi(\phi) = 2\pi((\phi/2\pi) - \lfloor\phi/2\pi\rfloor)$$

but Φ has nasty commutation relations. Another way to avoid mention of the phase operator is to talk about phase eigenstates. For example, one can define

$$|\phi\rangle = \sum_{n=0}^{\infty} e^{in\phi}|n\rangle.$$

These states are not normalizable, nor orthogonal, but they do provide a resolution of the identity,

$$\frac{1}{2\pi}\int_0^{2\pi} d\phi\, |\phi\rangle\langle\phi| = 1.$$

We can imperfectly formalize the idea of a "phase distribution" by projecting onto these states.

One could go on, but at the end of the day, these issues are really problems of the mind, not of reality. It does not matter that there does not exist a phase operator with all the formal properties we would like. In reality, experimentalists often perform measurements which are best described in English as "measuring the phase". Each of these experiments is certainly measuring *some* operator (though it might differ between experiments) and our actual task is to find what it is.

Note. A separate problem with the phase operator. Suppose one has defined a phase operator so that $[n, \phi] = i$. Then the expectation value of $[n, \phi]$ in any normalized state is i, but it is also 0 by the cyclic property of the trace.

However, this issue isn't particular to the phase operator. It also holds for position and momentum, or more generally any pair of conjugate variables. Here we've stumbled upon one of the (relatively few) cases where doing quantum mechanics without a proper functional analysis foundation yields clearly wrong results. Concretely, the problem is that sums like $\mathrm{tr}(n\phi)$ or more generally $\mathrm{tr}(xp)$ don't converge, so naively subtracting two such sums gives meaningless results. An overview of problems like these and their resolutions is given in *Mathematical surprises and Dirac's formalism in quantum mechanics*.

Note. It is important to remember that photons are defined as excitations of modes. So in situations where there are multiple relevant sets of modes, there are multiple relevant definitions of photons! Detectors that couple to different sets of these modes can register different numbers of photons, or even see photons when another detector sees vacuum. Bogoliubov transformations can be used to switch between the different mode descriptions.

In particle physics, this subtlety is usually glossed over because we work only in flat, free space, where the plane waves are the preferred set of modes. However, this breaks down for accelerating detectors, which detect photons corresponding to modes with different boundary conditions, and also in curved spacetime. These lead to the Unruh effect and Hawking radiation.

In applied physics (such as in cavity QED), this subtlety is usually not mentioned, even if there are nontrivial boundary conditions, because those boundary conditions are usually fixed. However, this breaks down when boundary conditions change. For example, if we begin with a monochromatic plane wave of photons of frequency ω, and then capture some by suddenly putting up walls to form a cavity, the cavity will contain photons in its modes, generally spanning a finite range of frequencies near ω. This seems paradoxical since it seems the photons must be "different", but putting up the walls ideally wouldn't create "new" photons. The resolution is that nothing has actually changed: the same state of the electromagnetic field is just being described in two different ways, using two different sets of mode functions. It is simultaneously true that the photons defined in terms of plane wave modes are monochromatic, and the photons defined in terms of cavity modes are not.

Before proceeding, it's useful to discuss some properties of coherent states, which were introduced in the notes on Undergraduate Physics.

- To establish notation, we recall that coherent states are defined as

$$a|\alpha\rangle = \alpha|\alpha\rangle, \quad |\alpha\rangle = e^{-|\alpha|^2/2} \sum_n \frac{z^n}{\sqrt{n!}} |n\rangle.$$

  The coherent states are not orthogonal,

$$\langle\beta|\alpha\rangle = \exp\left(-\frac{1}{2}|\beta-\alpha|^2 + \frac{1}{2}(\beta^*\alpha - \beta\alpha^*)\right), \quad |\langle\beta|\alpha\rangle|^2 = e^{-|\beta-\alpha|^2}.$$

However, they still satisfy a completeness relation,

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

- By inserting factors of the identity, we can expand states in terms of coherent states. This expansion has the odd property that a coherent state itself will get written as a combination of coherent states,

$$|\beta\rangle = \int \frac{d^2\alpha}{\pi} |\alpha\rangle\langle\alpha|\beta\rangle.$$

Thus in some sense $\langle\beta|\alpha\rangle$ plays the role of the Dirac delta; it is called a reproducing kernel.

- For an arbitrary state $|\psi\rangle$, the expansion coefficients are

$$\langle\alpha|\psi\rangle \equiv e^{-|\alpha|^2/2}\psi(\alpha^*), \quad \psi(z) = \sum_n \langle n|\psi\rangle \frac{z^n}{\sqrt{n!}}.$$

As an example, if $|\psi\rangle = |n\rangle$, then $\psi(z) = z^n/\sqrt{n!}$.

- For an operator $\hat{F}$, we can expand it in matrix elements as

$$\hat{F} = \sum_{mn} \hat{F}_{mn} |m\rangle\langle n|, \quad \hat{F}_{mn} = \langle m|\hat{F}|n\rangle.$$

We can do the same thing for coherent states, which yields

$$\hat{F} = \int \frac{d^2\beta}{\pi} \int \frac{d^2\alpha}{\pi} \exp\left(-\frac{|\beta|^2 + |\alpha|^2}{2}\right) F(\beta^*, \alpha) |\beta\rangle\langle\alpha|, \quad F(\beta^*, \alpha) = \sum_{mn} F_{mn} \frac{(\beta^*)^m \alpha^n}{\sqrt{m!n!}}.$$

- As yet another unintuitive result, the diagonal matrix elements $\langle\alpha|\hat{F}|\alpha\rangle$ completely determine $\hat{F}$, because

$$\langle\alpha|\hat{F}|\alpha\rangle e^{|\alpha|^2} = \sum_{mn} \frac{(\alpha^*)^m \alpha^n}{\sqrt{m!n!}} \hat{F}_{mn}.$$

Treating α and α^* as independent variables, we can recover $\hat{F}_{mn}$ by extracting the coefficient of $(\alpha^*)^m \alpha^n$ in the Taylor expansion of the left-hand side.

2.5 Quantum Correlation Functions

So far, we have only discussed correlation functions classically, instead of defining them starting from a fully quantum framework. We do this here, gaining new insight in the process.

- Let us model an absorbing detector as an atom that couples to the electromagnetic field and thereby absorbs light. For example, in the dipole approximation, the interaction is $-\mathbf{d} \cdot \mathbf{E}(\mathbf{r}, t)$ where $\mathbf{r}$ is the location of the atom.

- In interaction picture, and standard AMO normalization conventions,

$$\mathbf{E}(\mathbf{r}, t) = i \sum_{\mathbf{k},s} \sqrt{\frac{\hbar\omega_k}{2\epsilon_0 V}} \, \mathbf{e}_{\mathbf{k},s} \left(a_{\mathbf{k},s} e^{i(\mathbf{k}\cdot\mathbf{r} - \omega t)} - a^\dagger_{\mathbf{k},s} e^{-i(\mathbf{k}\cdot\mathbf{r} - \omega t)} \right)$$

where $\mathbf{e}_{\mathbf{k},s}$ is a polarization vector.

- We can decompose the field as

$$\mathbf{E}(\mathbf{r},t) = \mathbf{E}^+(\mathbf{r},t) + \mathbf{E}^-(\mathbf{r},t)$$

where the components are called the positive and negative frequency parts, and contain annihilation and creation operators, respectively; they are Hermitian adjoints of each other.

- In a photodetection event, the atom is excited from state $|g\rangle$ to $|e\rangle$, while the field goes from state $|i\rangle$ to $|f\rangle$. Let us suppose for concreteness that the atom is coupled to the field for a very short time window around t. Then the probability for a transition is given by first-order perturbation theory,

$$P \propto \sum_f |\langle e|\mathbf{d}|g\rangle|^2 |\langle f|\mathbf{E}^+|i\rangle|^2 \propto \sum_f |\langle f|\mathbf{E}^+|i\rangle|^2.$$

Here $|f\rangle$ ranges over all possible states of the field, so we have

$$P \propto \sum_f \langle i|\mathbf{E}^-|f\rangle\langle f|\mathbf{E}^+|i\rangle = \langle i|\mathbf{E}^- \cdot \mathbf{E}^+|i\rangle.$$

- More generally, the initial state is described by a density matrix ρ, so

$$P \propto \mathrm{tr}(\rho \mathbf{E}^- \cdot \mathbf{E}^+).$$

Thus, when we measure the intensity of light at a point $x = (\mathbf{r}, t)$, we are really measuring

$$G^{(1)}(x,x) = \mathrm{tr}\left(\rho E^-(x) E^+(x)\right)$$

where we have suppressed polarizations. Note that this expression is normal ordered, which is due to the fact that we used an absorbing detector.

- More generally, we can define

$$G^{(1)}(x_1,x_2) = \mathrm{tr}\left(\rho E^-(x_1) E^+(x_2)\right).$$

Just as in the classical case, this is relevant if we arrange for the field at a point to be related to the fields at two different points,

$$E^+(x) = K_1 E^+(x_1) + K_2 E^+(x_2)$$

by interferometry, as will be justified below. The first-order correlation function is

$$g^{(1)}(x_1,x_2) = \frac{G^{(1)}(x_1,x_2)}{\sqrt{G^{(1)}(x_1,x_1)G^{(1)}(x_2,x_2)}}.$$

It satisfies the same bounds as the classical version. The only minor difference is that we are now working with a spacetime argument rather than a temporal one, just because the former was more convenient starting from quantum fields.

- For example, suppose the initial state $|n\rangle$ contains n excitations for some mode, and none for all other modes. The only terms that contribute in the correlation function are those with creation and annihilation operators for that mode, so we lose the sum over modes. The only relevant part of the field operator is

$$\mathbf{E}^+(x) \sim \hat{a} e^{i(\mathbf{k}\cdot\mathbf{r}-\omega t)}$$

  where $\hat{a}$ is the annihilation operator for the occupied mode. We then have

$$\begin{aligned} G^{(1)}(x_1, x_2) &= \langle n|E^-(x_1)E^+(x_2)|n\rangle \\ &\propto \langle n|a^\dagger a|n\rangle e^{i(\mathbf{k}\cdot(\mathbf{r}_1-\mathbf{r}_2)-\omega(t_1-t_2))} \\ &= n e^{ik\cdot(x_1-x_2)}. \end{aligned}$$

  Then we have $|g^{(1)}(x_1, x_2)| = 1$ as expected.

- As another example, suppose we begin with a coherent state $|\alpha\rangle$ in one mode. Then

$$G^{(1)}(x_1, x_2) = \langle\alpha|a^\dagger a|\alpha\rangle e^{ik\cdot(x_1-x_2)} = |\alpha|^2 e^{ik\cdot(x_1-x_2)}$$

  which again gives $|g^{(1)}(x_1, x_2)| = 1$. Evidently, the difference between number eigenstates and coherent states only appears in higher correlation functions. For perfect first-order coherence, it suffices to have any pure state that only contains excitations in one mode.

We now discuss how an interferometry experiment works, quantum mechanically.

- It is clear that classically, an interference experiment can be thought of as measuring the intensity of the field

$$E^+(x) = K_1 E^+(x_1) + K_2 E^+(x_2),$$

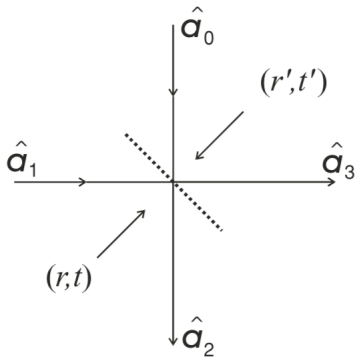
  and above we simply assumed this held quantum mechanically. This is essentially correct, but it is not obvious how to make this idea precise. In particular, naive application of this equation when photon numbers are low can lead to paradoxes.

- To warm up, we consider a beam splitter with classical reflection and transmission coefficients r and t. The input field E_1 is then split into fields at the two outputs,

$$E_2 = rE_1, \quad E_3 = tE_1.$$

  At the quantum level, these classical quantities should be replaced with the annihilation operators for the corresponding modes. But it is invalid to write, say, $a_2 = ra_1$ because this would not satisfy the harmonic oscillator commutation relations.

- The problem is that we are doing something inherently nonunitary. Since the final states comprise two modes, the initial states should as well. There is a second, unoccupied field mode that we need to describe the input, which corresponds to approaching the beam splitter from the other direction.



Let the reflection and transmission coefficients from this side be r' and t'. Then

$$\begin{pmatrix} a_2 \\ a_3 \end{pmatrix} = \begin{pmatrix} t' & r \\ r' & t \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \end{pmatrix}.$$

- The harmonic oscillator commutation relations hold if

$$|r'| = |r|, \quad |t'| = |t|, \quad |r|^2 + |t|^2 = 1, \quad r^* t' + r' t^* = 0, \quad r^* t + r' t'^* = 0.$$

These "reciprocity relations" can also be derived classically on the basis of energy conservation.

We now give some examples with beam splitters.

- In a 50/50 beam splitter corresponding to a single dielectric layer,

$$a_2 = \frac{1}{\sqrt{2}}(a_0 + i a_1), \quad a_3 = \frac{1}{\sqrt{2}}(i a_0 + a_1).$$

This is just like how the S-matrix works in quantum field theory; in that case as well, we have "in" and "out" modes/states, and we want to find the transformation that relates them.

- Continuing the example, consider the input state with n photons in mode 1. This can be written in terms of the "out" modes 2 and 3 as

$$|0\rangle_0 |n\rangle_1 = \frac{(a_1^\dagger)^n}{\sqrt{n!}} |\Omega\rangle = \frac{1}{\sqrt{n!}} \left(\frac{i a_2^\dagger + a_3^\dagger}{\sqrt{2}} \right)^n |\Omega\rangle.$$

Note that there is no notion of time dependence. This is just as in classical scattering theory and in quantum field theory, and it is because in all cases we are working with "in" and "out" modes which are eigenstates of the free Hamiltonian. As always, we can recover an intuitive time dependence using wavepackets.

- When $n = 1$, the final state is simply

$$\frac{1}{\sqrt{2}} \left(i|1\rangle_2 |0\rangle_3 + |0\rangle_2 |1\rangle_3 \right)$$

just as we would naively expect, and as we asserted earlier. Note that this is an entangled state of the photon field; beam splitters generically produce entanglement.

- For general n, the amplitude for m and $n-m$ photons in the final state is

$$(\langle m|_2\langle n-m|_3)\,(|0\rangle_0|n\rangle_1) \propto \frac{1}{\sqrt{n!}}\sqrt{m!}\sqrt{(n-m)!}\binom{n}{m} = \binom{n}{m}^{1/2}.$$

  In particular, the probability is proportional to $\binom{n}{m}$. Therefore, we get precisely the same result as if we naively assumed that each photon individually had a 50/50 chance of going in each direction, which we used in a previous example.

- The reason is that two mistakes cancel out. In the naive argument, the photons are treated as distinguishable when they actually aren't, overcounting states where the photons are split more evenly between the final modes. But the argument also ignores the fact that bosons have a higher amplitude to bunch up. The lesson here is that we should commit to either the naive classical picture or the fully quantum one; adding ad hoc corrections to go from the former to the latter is more difficult than just doing everything quantum mechanically from the start.

- As another example, consider an initial coherent state,

$$|0\rangle_0|\alpha\rangle_1 = \exp\left(\alpha a_1^\dagger - \alpha^* a_1\right)|\Omega\rangle = \exp\left(\frac{\alpha}{\sqrt{2}}(ia_2^\dagger + a_3^\dagger) - \frac{\alpha^*}{\sqrt{2}}(-ia_2 + a_3)\right)|\Omega\rangle.$$

  Since the operators for modes 2 and 3 commute, we can factorize this as

$$|0\rangle_0|\alpha\rangle_1 = \exp\left(\frac{i\alpha}{\sqrt{2}}a_2^\dagger - \frac{(-i)\alpha^*}{\sqrt{2}}a_2\right)\exp\left(\frac{\alpha}{\sqrt{2}}a_3^\dagger - \frac{\alpha^*}{\sqrt{2}}a_3\right)|\Omega\rangle = \left|\frac{i\alpha}{\sqrt{2}}\right\rangle_2\left|\frac{\alpha}{\sqrt{2}}\right\rangle_3.$$

  That is, a coherent state becomes a product of coherent states, and there is no entanglement between the output modes. This is another manifestation of how coherent states behave essentially classically.

- This result can also be understood in the naive classical picture. Recall that the photons in a coherent state are Poisson distributed. Each photon hitting the beam splitter chooses randomly and independently. Hence photon arrival at each output port is also a Poisson process, so the output states are coherent. Furthermore, if one restricts to final states with a definite number of "out" photons, they are divided between the "out" modes by a binomial distribution. (Of course, this classical picture ignores phase information, and so is useless when considering interferometry.)

- This result also shows, once again, how true single-photon sources differ from coherent states with low α. Beam splitting the latter never produces any entanglement, while beam splitting the former produces maximal entanglement.

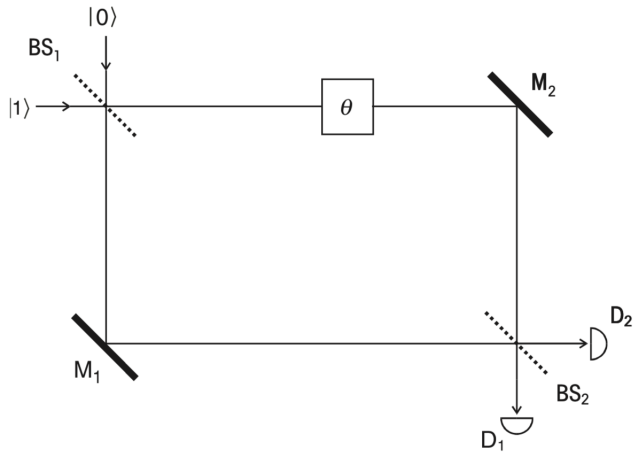
- Interferometry experiments can be described by recombining the beams with another beam splitter, as in a Mach–Zehnder interferometer, or more simply by having photons in both "in" modes. For example, we have

$$|1\rangle_0|1\rangle_1 = \frac{1}{2}(a_2^\dagger + ia_3^\dagger)(ia_2^\dagger + a_3^\dagger)|\Omega\rangle = \frac{i}{2}\left(|2\rangle_2|0\rangle_3 + |0\rangle_2|2\rangle_3\right).$$

  The amplitude for having one photon in each "out" mode is zero, by destructive interference.

- This is known as the Hong–Ou–Mandel effect, and it is interesting because it was the first direct measurement of interference between two photons. Of course, we know that classically, the visibility of interference effects depends on coherence. In the quantum mechanical formalism, this is accounted for by replacing the states $|1\rangle_0$ and $|1\rangle_1$ with appropriate wavepackets. For example, one can show that if the wavepackets hit the beam splitter at different times, no interference occurs, because the photons are distinguishable by their timing.
- Our formalism can also be applied to interference in the double slit experiment. This is more complicated because there is more spatial structure, but it isn't different in principle. For simplicity, we focus on beam splitters below.

Example. The Mach–Zehnder interferometer. We consider two 50/50 beam splitters with an adjustable phase shift θ inside the interferometer.



Label the pairs of modes 0/1, 2/3, 4/5, and 6/7. If a single photon enters, the state is

$$|\psi\rangle = |0\rangle_0|1\rangle_1 = \frac{1}{\sqrt{2}}(|01\rangle + i|10\rangle)_{23}$$

where we switch to an abbreviated notation for convenience. After the phase shifter and mirror,

$$|\psi\rangle = \frac{1}{\sqrt{2}}(e^{i\theta}|01\rangle + i|10\rangle)_{45}.$$

After the second beam splitter, we have

$$|\psi\rangle = \frac{1}{2}((e^{i\theta} - 1)|01\rangle + i(e^{i\theta} + 1))_{67}.$$

This is just what we would expect for a Mach–Zehnder interferometer.

The subscripts can get annoying, which is why it is conventional to drop them. Instead of thinking of the pairs of modes as different bases in a Hilbert space, we think of the devices as implementing transformations on that Hilbert space, e.g. the first step would be written as

$$|01\rangle \to \frac{1}{\sqrt{2}}(|01\rangle + i|10\rangle).$$

To demonstrate this notation, we repeat the derivation for a coherent state. This is necessary because our intuition for what a Mach–Zehnder interferometer does comes from classical physics,

which corresponds to coherent states. We have

$$|0,\alpha\rangle \to \left|\frac{i\alpha}{\sqrt{2}},\frac{\alpha}{\sqrt{2}}\right\rangle \to \left|\frac{ie^{i\theta}\alpha}{\sqrt{2}},\frac{\alpha}{\sqrt{2}}\right\rangle \to \left|\frac{i(e^{i\theta}+1)\alpha}{\sqrt{2}},\frac{(e^{i\theta}-1)\alpha}{\sqrt{2}}\right\rangle.$$

The ratio of intensities coming out the final beam splitter is the same, but the derivation looks completely different: the "interference" occurs in the complex coherent state parameters, rather than in the quantum amplitudes. Of course, at the end of the day everything is described by quantum amplitudes, but this example shows how a more classical-looking kind of interference can emerge from that.

Example. The Elitzur–Vaidman bomb detector. Suppose we configure a Mach–Zehnder interferometer with two detectors in the final arms, such that when single photons are inserted, the detector D_1 always fires and the detector D_2 never does, due to complete destructive interference. Suppose somebody puts a bomb in one arm of the interferometer, which explodes whenever a photon hits it.

The bomb effectively performs a position measurement, collapsing the superposition. It explodes with a 50% chance. The other 50% of the time, the photon collapses onto the other arm, then hits the beam splitter alone. Since there is not destructive interference in this case, the detector D_2 fires 25% of the time.

This thought experiment really becomes puzzling once you ask further questions. What we learn is that 50% of the time, the photon wasn't in the arm with the bomb. And in those cases, 50% of the time, detector D_2 fires. Therefore, in those cases, the photon can tell that a bomb is in an arm of the interferometer even if the photon was "never there". According to science popularizers and AMO papers, this proves that in quantum mechanics, photons can teleport instantaneously, and do not obey causality. In related experiments such as the delayed-choice quantum eraser, one can even hear claims that photons can retroactively change the past, or move between parallel timelines or universes, or literally break *the rules of logic themselves*, requiring a theory of "quantum logic".

Upon reflection, anybody who knows how quantum mechanics works will see that these mind-blowing statements are all completely wrong. What the thought experiment really tells us is a familiar lesson from every first quantum mechanics class. The photon is not in one arm or the other, or both, or neither, it's in a *superposition* of the two, and superpositions obey fundamentally different rules. Ignoring this and pretending that the photon is a classical bullet in one path or the other, but that we simply don't know which, leads to paradoxical results, as was already found by EPR almost a century ago. "Mindblowing" but incorrect conclusions come from trying to use classical intuition when it does not apply, not from the quantum mechanics itself, which is perfectly straightforward, deterministic, and causal. Such statements are repeated because everybody expects to hear that quantum mechanics makes no sense, while the actual point is the opposite: quantum mechanics does make sense, and what doesn't make sense is modeling our quantum world classically.

Finally, we discuss higher-order correlation functions.

- As we've seen, the first-order correlation function is generally insensitive to photon number statistics within one mode, not being able to distinguish a coherent state from a number state. Instead, we need to use the second-order correlation function.

- We model our absorbing detectors as atoms, as in our discussion of first-order correlation functions. Suppose that the detectors are at positions $\mathbf{r}_1$ and $\mathbf{r}_2$, and are turned on for short windows around t_1 and t_2. Then the probability for both to transition is, at lowest order in

perturbation theory,

$$P \propto \sum_f |\langle f|\mathbf{E}^+(x_2)\mathbf{E}^+(x_1)|i\rangle|^2$$

and performing the same manipulations and suppressing vector indices gives

$$P \propto \mathrm{tr}(\rho E^-(x_1)E^-(x_2)E^+(x_2)E^+(x_1)) \equiv G^{(2)}(x_1, x_2)$$

where we have used the fact that the E^+ operators commute.

- We are hence motivated to define the second-order coherence function as

$$g^{(2)}(x_1, x_2) = \frac{G^{(2)}(x_1, x_2)}{G^{(1)}(x_1, x_1)G^{(1)}(x_2, x_2)}.$$

A quantum state is second-order coherent if $|g^{(2)}(x_1, x_2)| = 1$. Similarly one can define $g^{(2)}(\tau)$.

- Note that the analogous definition to $g^{(1)}(x_1, x_2)$ would have four arguments, and involve

$$G^{(2)}(x_1, x_2, x_3, x_4) \equiv \mathrm{tr}(\rho E^-(x_1)E^-(x_2)E^+(x_3)E^+(x_4)).$$

We are using a less general definition because it is what is relevant in practice.

- As an example, consider any state with only a single mode occupied. Then

$$g^{(2)}(\tau) = \frac{\langle a^\dagger a^\dagger aa\rangle}{\langle a^\dagger a\rangle^2} = \frac{\langle n(n-1)\rangle}{\langle n\rangle^2} = 1 + \frac{(\mathrm{var}\, n) - \langle n\rangle}{\langle n\rangle^2}.$$

Then in a single mode coherent state, $g^{(2)}(\tau) = 1$. In a single mode thermal state, which is exponentially distributed, $g^{(2)}(\tau) = 2$. In a single mode number eigenstate, $g^{(2)}(\tau) = 1 - 1/n$, as we derived more laboriously earlier. Of course, nontrivial time dependence, which is needed to discuss bunching, only enters when we consider occupancy of multiple modes.

- Earlier, we noted that several sources of chaotic light had $g^{(2)}(0) = 2$. This holds quite generally, and it is because in such cases one has a Gaussian field distribution and hence can use Wick's theorem to expand the four-point correlation function, giving

$$g^{(2)}(\tau) = 1 + |g^{(1)}(\tau)|^2.$$

This shows that $g^{(2)}(0) = 2$ for any chaotic light.

- Similarly one can define an n^{th} order correlation function,

$$g^{(n)}(x_1, \ldots, x_n) = \frac{G^{(n)}(x_1, \ldots, x_n)}{G^{(1)}(x_1, x_1)\ldots G^{(1)}(x_n, x_n)}$$

where

$$G^{(n)}(x_1, \ldots, x_n) = \mathrm{tr}\left(\rho E^-(x_1), \ldots, E^-(x_n)E^+(x_n)\ldots E^+(x_1)\right).$$

Just like the general four-argument second-order coherence function, these correlation functions are perfectly mathematically sensible, but they are less important in practice. They would describe correlations in experiments involving n detectors.

- We say a state is n^{th} order coherent if

$$|g^{(n)}(x_1, \ldots, x_n)| = 1.$$

This holds if the correlation function factors,

$$G^{(n)}(x_1, \ldots, x_n) = G^{(1)}(x_1, x_1) \ldots G^{(1)}(x_n, x_n)$$

which is true for coherent states. For single mode states, these higher correlation functions describe higher moments of the photon number distribution.

- Again, one can define more general correlation functions with $2n$ arguments. Even more generally, one can work with other operator orderings, such as symmetric ordering or anti-normal ordering, which can be used to describe exotic experiments.

3 Quantum States of Light

3.1 Phase Space Distributions

In this section, we discuss more exotic states of light in detail. It is useful to visualize these states using phase space (quasi)probability distributions.

- It is natural to pass from pure states to mixed states, since the latter are already explicitly probability distributions anyway.
- It turns out that any density operator can be represented uniquely in the form

$$\hat{\rho} = \int P(\alpha)|\alpha\rangle\langle\alpha| \, d^2\alpha$$

  where $P(\alpha)$ is called the Glauber–Sudarshan P function, which is thought of as a distribution over phase space. To verify that it is normalized, note that

$$\operatorname{tr}\hat{\rho} = \int \sum_n P(\alpha)\langle n|\alpha\rangle\langle\alpha|n\rangle \, d^2\alpha = \int \sum_n P(\alpha)\langle\alpha|n\rangle\langle n|\alpha\rangle \, d^2\alpha = \int P(\alpha) \, d^2\alpha = 1.$$

- Unlike a standard probability distribution, the P function can be negative, or more singular than a Dirac delta function. States that don't have either property are defined to be "classical", in the sense that they have semiclassical analogues. For example, thermal states are classical under this definition, while squeezed states aren't.
- This is not a very restrictive condition: it turns out that *all* pure states besides the coherent states are "nonclassical". Yet in practice, it is nontrivial to make significantly nonclassical states in the lab. This explains why the semiclassical AMO intuition of "quantum fluctuations" works so well: for many purposes, we really can think of a quantum state as a classical state with overlaid stochastic fluctuations.
- The definition of $P(\alpha)$ is easy to use, but it doesn't directly tell us how to compute it. To do so, we note that

$$\langle -u|\hat{\rho}|u\rangle = e^{-|u|^2} \int P(\alpha) e^{-|\alpha|^2} e^{\alpha^* u - \alpha u^*} \, d^2\alpha$$

  where the $|\pm u\rangle$ are coherent states. This can be rewritten as

$$\left(\langle -u|\hat{\rho}|u\rangle e^{|u|^2}\right) = \int \left(P(\alpha)e^{-|\alpha|^2}\right) e^{\alpha^* u - \alpha u^*} \, d^2\alpha.$$

  This has the form of a two-dimensional Fourier transform, where the conjugate pairs are $\operatorname{Re}\alpha$ and $\operatorname{Im} u$, and $\operatorname{Im}\alpha$ and $-\operatorname{Re} u$. By inverting the Fourier transform, we thus find

$$P(\alpha) = \frac{e^{|\alpha|^2}}{\pi^2} \int e^{|u|^2} \langle -u|\hat{\rho}|u\rangle e^{u^*\alpha - u\alpha^*} \, d^2u.$$

- For example, for a pure coherent state $\hat{\rho} = |\beta\rangle\langle\beta|$, this formula gives

$$P(\alpha) = e^{|\alpha|^2} e^{-|\beta|^2} \frac{1}{\pi^2} \int e^{u^*(\alpha-\beta) - u(\alpha^*-\beta^*)} \, d^2u.$$

Expanding in components shows that the integral is a representation of a two-dimensional Dirac delta function, which gives

$$P(\alpha) = \delta(\alpha - \beta).$$

Thus the coherent states are "classical".

- As another example, for a pure number state $\hat{\rho} = |n\rangle\langle n|$, we have

$$P(\alpha) = \frac{e^{|\alpha|^2}}{n!}\frac{1}{\pi^2}\int (-u^* u)^n e^{u^*\alpha - u\alpha^*}\, d^2u.$$

This integral seems to diverge, but note that it can be written as a derivative of a delta function,

$$P(\alpha) = \frac{e^{|\alpha|^2}}{n!}\frac{\partial^{2n}}{\partial\alpha^n \partial\alpha^{*n}}\frac{1}{\pi^2}\int e^{u^*\alpha - u\alpha^*}\, d^2u = \frac{e^{|\alpha|^2}}{n!}\frac{\partial^{2n}}{\partial\alpha^n \partial\alpha^{*n}}\delta(\alpha).$$

Thus the number states are not "classical".

- Neither of these results are in accord with the intuitive phase space distributions we mentioned above. However, this is acceptable because there are actually many different ways to define phase space distributions, and we'll see some more intuitive ones shortly.

- Classically, phase space functions can be written in the form $G(a, a^\dagger)$. In quantum mechanics, these become operators, and we must specify the ordering. In normal ordering, we have

$$\hat{G}^{(N)}(a, a^\dagger) = \sum_{nm} C_{nm} (a^\dagger)^n a^m$$

where we suppress hats on a and $a^\dagger$ for brevity.

- We can compute the expectation value of $\hat{G}^{(N)}$ using the P function as we would expect,

$$\begin{aligned}
\langle \hat{G}^{(N)}(a, a^\dagger)\rangle &= \mathrm{tr}\int P(\alpha)\sum_{nm} C_{nm}(a^\dagger)^n a^m |\alpha\rangle\langle\alpha|\, d^2\alpha \\
&= \int P(\alpha)\sum_{nm} C_{nm}\langle\alpha|(a^\dagger)^n a^m|\alpha\rangle\, d^2\alpha \\
&= \int P(\alpha)\sum_{nm} C_{nm}(\alpha^*)^n \alpha^m\, d^2\alpha \\
&= \int P(\alpha) G^{(N)}(\alpha, \alpha^*)\, d^2\alpha.
\end{aligned}$$

This is the optical coherence theorem: the expectation value of a normally ordered operator is the P function weighted average. As such, the correlation functions defined above can be evaluated in terms of P functions, and the results can differ from the classical expectation precisely because P can be negative.

We now motivate some more phase space distributions.

- Under certain conditions, operators besides $\hat{\rho}$ can be given P representations,

$$\hat{B} = \int B_p(\alpha, \alpha^*)|\alpha\rangle\langle\alpha|\, d^2\alpha.$$

Then the expectation value is

$$\langle\hat{B}\rangle = \mathrm{tr}(\hat{B}\hat{\rho}) = \int B_p(\alpha, \alpha^*)\,\mathrm{tr}\,(|\alpha\rangle\langle\alpha|\hat{\rho})\; d^2\alpha = \int B_p(\alpha, \alpha^*)\langle\alpha|\hat{\rho}|\alpha\rangle\, d^2\alpha.$$

- This motivates the definition of the Husimi Q function,

$$Q(\alpha) = \frac{1}{\pi}\langle\alpha|\hat{\rho}|\alpha\rangle$$

where the $1/\pi$ is included for normalization. Unlike the P function, the Q function is always positive, so it is a genuine probability distribution. It is easier to compute, but harder to use.

- Now, we could also write $\hat{B}$ in terms of a Q representation,

$$B_Q(\alpha, \alpha^*) = \langle\alpha|\hat{B}|\alpha\rangle$$

in which case expectation values naturally involve the P function,

$$\langle\hat{B}\rangle = \operatorname{tr}(\hat{B}\hat{\rho}) = \int P(\alpha)\operatorname{tr}(\hat{B}|\alpha\rangle\langle\alpha|)\,d^2\alpha = \int P(\alpha)B_Q(\alpha,\alpha^*)\,d^2\alpha.$$

Thus in some sense the P representation and Q representation are dual.

- The earliest introduced phase space distribution is the Wigner function,

$$W(q,p) = \frac{1}{2\pi\hbar}\int\langle q + x/2|\hat{\rho}|q - x/2\rangle e^{ipx/\hbar}\,dx$$

where the states $|q \pm x/2\rangle$ are position eigenstates. For a pure state $\hat{\rho} = |\psi\rangle\langle\psi|$,

$$W(q,p) = \frac{1}{2\pi\hbar}\int\psi^*(q - x/2)\psi(q + x/2)e^{ipx/\hbar}\,dx.$$

- The point of the Wigner function is that it marginalizes to the position and momentum distributions. Some routine computations show that

$$\int W(q,p)\,dp = |\psi(q)|^2, \quad \int W(q,p)\,dq = |\varphi(p)|^2.$$

The Wigner function can take on negative values. It turns out that negative values imply a nonclassical state, but some nonclassical states have a positive Wigner function.

- It can be shown that the Q function satisfies the optical coherence theorem for anti-normally ordered operators. Moreover, the Wigner function satisfies the optical coherence theorem for Weyl, or symmetrically ordered functions of p and q (suppressing hats),

$$\langle\hat{G}(q,p)^{(W)}\rangle = \int W(q,p)G^{(W)}(q,p)\,dqdp$$

where, for example, the Weyl ordering of qp is $(qp + pq)/2$. Hence the many choices of phase space distribution just reflect the many choices of operator ordering prescription.

Note. The Q function can be thought of as a smoothed version of the Wigner function, which is a smoothed version of the P function; more technically, the smoothing is done by Gaussian convolutions. For example, for coherent states, $Q(\alpha)$ and $W(x,p)$ are both Gaussians, while $P(\alpha)$ is a delta function at the origin. For a number state, $Q(\alpha)$ is peaked at the appropriate radius, while $W(x,p)$ oscillates between positive and negative values at smaller radii, and $P(\alpha)$ is extremely singular at the origin. For a "cat" state formed by superposing two coherent states, $Q(\alpha)$ has two peaks, $W(x,p)$ has rapid positive/negative oscillations between the peaks, and $P(\alpha)$ is again extremely singular. For thermal states, all three are Gaussians centered at the origin, with slightly different widths.

Like ordinary probability distributions, these phase space distributions can be described in terms of characteristic functions.

- A classical probability distribution $\rho(x)$ has moments

$$\langle x^n \rangle = \int dx\, x^n \rho(x)$$

which determine the probability distribution. To see this, introduce the characteristic function,

$$C(k) = \langle e^{ikx} \rangle = \int dx\, e^{ikx} \rho(x) = \sum_n \frac{(ik)^n}{n!} \langle x^n \rangle.$$

Then the probability distribution is just the Fourier transform of the characteristic function,

$$\rho(x) = \frac{1}{2\pi} \int dk\, e^{-ikx} C(k)$$

which is in turn determined by the moments. We can also recover the moments from the characteristic function,

$$\langle x^n \rangle = \frac{1}{i^n} \frac{d^n C(k)}{dk^n}\bigg|_{k=0}.$$

- Similarly, we can define quantum characteristic functions, where the equivalent of k now ranges over two dimensions. However, we again run into ordering issues. Three simple options are

$$C_W(\lambda) = \mathrm{tr}(\rho e^{\lambda a^\dagger - \lambda^* a}), \quad C_N(\lambda) = \mathrm{tr}(\rho e^{\lambda a^\dagger} e^{-\lambda^* a}), \quad C_A(\lambda) = \mathrm{tr}(\rho e^{-\lambda a^\dagger} e^{\lambda^* a})$$

which correspond to Weyl, normal, and antinormal ordering; then the expectation values of Weyl, normal, and antinormal ordered operators, respectively, can be recovered by differentiation.

- These three functions are simply related to each other. By Glauber's theorem,

$$C_W(\lambda) = C_N(\lambda) e^{-|\lambda|^2/2} = C_A(\lambda) e^{|\lambda|^2/2}.$$

This motivates defining a one-parameter family of characteristic functions,

$$C(\lambda, s) = \mathrm{tr}(\rho e^{\lambda a^\dagger - \lambda^* a + s|\lambda|^2/2})$$

where s parametrizes the amount of "smoothing".

- As expected, these characteristic functions are related to the phase space probability distributions by Fourier transforms. For example, for the Q function,

$$C_A(\lambda) = \mathrm{tr}(e^{\lambda a^\dagger} \rho e^{-\lambda^* a}) = \frac{1}{\pi} \int d^2\alpha\, \langle \alpha | e^{\lambda a^\dagger} \rho e^{-\lambda^* a} | \alpha \rangle = \int d^2\alpha\, Q(\alpha) e^{\lambda \alpha^* - \lambda^* \alpha}.$$

Similarly, for the P function, we note that

$$C_N(\lambda) = \mathrm{tr}(\rho e^{\lambda a^\dagger} e^{-\lambda^* a}) = \int d^2\alpha\, P(\alpha) \langle \alpha | e^{\lambda a^\dagger} e^{-\lambda^* a} | \alpha \rangle = \int d^2\alpha\, P(\alpha) e^{\lambda \alpha^* - \lambda^* \alpha}.$$

The same holds for the Wigner function, with the appropriate conversion of $\alpha = (q + ip)/\sqrt{2}$.

In some cases, it is easier to calculate a phase space distribution by calculating the characteristic function first.

3.2 Squeezed States

We begin with the basics of quadrature squeezing.

- We define the quadrature operators X_1 and X_2 as above, so the uncertainty relation is

$$(\Delta X_1)(\Delta X_2) \geq \frac{1}{4}.$$

  A state is said to be squeezed if $(\Delta X_i)^2 < 1/4$ for either i. Some, but not all squeezed states saturate the uncertainty relation. It can be shown that squeezed states must have a P-function that is negative in some part of phase space.

- As a concrete example of a squeezed state, note that the scale of the quadrature operators for a harmonic oscillator depends on the parameters m and ω. Thus, the vacuum state for one oscillator is squeezed with respect to an oscillator with different parameters, so a squeezed state can be produced by suddenly changing these operators.

- Specifically, let us define

$$e^\gamma = \sqrt{\frac{m_1\omega_1}{m_2\omega_2}}.$$

  Then the creation and annihilation operators of the two harmonic oscillators are related by

$$a_1 + a_1^\dagger = e^\gamma(a_2 + a_2^\dagger), \quad a_1 - a_1^\dagger = e^{-\gamma}(a_2 - a_2^\dagger)$$

  because the oscillators have the same x and p. Solving, we find

$$a_2 = a_1 \cosh\gamma - a_1^\dagger \sinh\gamma, \quad a_2^\dagger = -a_1 \sinh\gamma + a_1^\dagger \cosh\gamma$$

  which is the standard form of a Bogoliubov transformation, i.e. a relation between different sets of creation and annihilation operators.

- The original vacuum state satisfies

$$a_1|0\rangle_{(1)} = (a_2 \cosh\gamma + a_2^\dagger \sinh\gamma)|0\rangle_{(1)} = 0.$$

  Since this vacuum state is even in x, only states of the form $|2n\rangle_{(2)}$ should be present in $|0\rangle_{(1)}$. The coefficients of these states are related to each other by the constraint above, and solving the resulting recurrence relation gives

$$|0\rangle_{(1)} = \frac{1}{\sqrt{\cosh\gamma}} \exp\left(-\frac{1}{2}\tanh\gamma\, a_2^\dagger a_2^\dagger\right)|0\rangle_{(2)}.$$

  This state has ΔX_1 squeezed by a factor of $e^{-\gamma}$. In the limit of infinite squeezing, we recover a position eigenstate.

- To generalize this construction, we define the unitary "squeezing operator",

$$S(\xi) = \exp\left(\frac{1}{2}(\xi^* a^2 - \xi a^{\dagger 2})\right), \quad \xi = re^{i\theta}$$

  which looks very similar to the translation operator $D(\alpha)$, but with a^2 instead of a. Acting with the squeezing operator on the vacuum gives a squeezed state, by our logic above where $e^{-\gamma} = e^{-r}/2$. The direction of squeezing can be varied with θ, where the X_2 quadrature is squeezed for $\theta = \pi$.

- A general family of squeezed states can be constructed by applying a squeezing operator to the vacuum, followed by a translation. **(todo)**

We now discuss the production and detection of squeezed light.

-
-
-

3.3 Other Nonclassical States

4 Radiative Transitions

4.1 Rabi Oscillations

- Previously, we have discussed the interaction of atoms with a quantized radiation field. However, this analysis used Fermi's golden rule, which implicitly requires first-order time-dependent perturbation theory to hold.
- This is a fragile assumption; for instance, it breaks down if we just wait for a longer time. For spontaneous emission, the transition rate clearly cannot remain constant for all time, because there are a finite number of transitions that can happen. On the other hand, applying higher-order time-dependent perturbation theory is painful.
- The so-called "Wigner–Weisskopf" theory of spontaneous emission in free space addresses this problem. It uses physically motivated approximations to arrive at the intuitive conclusion that the amplitude to be in the excited state falls exponentially over time, and furthermore remains zero afterward. This is because the atom is coupled to a continuum of modes; the amplitude "gets lost" among all of these modes and never comes back.
- Now consider the distinct situation where an atom is placed in a monochromatic field, which induces absorption and stimulated emission. For simplicity, we treat the atom as a two-state system, with energies E_0 and E_1 and $E_1 - E_0 = \hbar\omega$.
- If the field has a high occupancy, this situation can be treated semiclassically. We investigated this situation in the notes on Undergraduate Physics. Let the driving have the form

  $$H_1(t) = (-\mathbf{d}\cdot\mathbf{E}_0)\cos(\omega_0 t) = -\frac{\mathbf{d}\cdot\mathbf{E}_0}{2}(e^{i\omega_0 t} + e^{-i\omega_0 t}).$$

  The rotating wave approximation amounts to ignoring the negative-frequency part,

  $$H_1(t) \approx -\frac{\mathbf{d}\cdot\mathbf{E}_0}{2}\, e^{-i\omega_0 t}.$$

  The name comes from NMR, where it corresponds to ignoring the part of the applied magnetic field that does not rotate in the same sense as the precessing magnetic moments.
- We define $\mathcal{M} = \langle 1|(-\mathbf{d}\cdot\mathbf{E}_0)|0\rangle$. For convenience, let $\Delta = \omega - \omega_0$. Working in interaction picture, we let

  $$|\psi(t)\rangle = c_0(t)|0\rangle + c_1(t)|1\rangle.$$

  Using first-order perturbation theory and an initial state $|0\rangle$, we have

  $$c_1(t) = \frac{\mathcal{M}}{2i\hbar}\int_0^t dt'\, e^{i\Delta t'} = \frac{\mathcal{M}}{2\hbar\Delta}\left(1 - e^{i\Delta t}\right)$$

  which gives an oscillatory probability of

  $$P_1(t) = \frac{|\mathcal{M}|^2}{\hbar^2}\frac{\sin^2(\Delta t/2)}{\Delta^2}.$$

- Again, one might ask how this fits with the naive picture of absorption; we expect there should be a regime where the probability increases linearly with time. As usual, this result is obtained if the driving involves a continuum of states, as then the logic of Fermi's golden rule applies. In a proper quantum treatment, these would be the states of the photon field, as realistically the incoming light has photons occupying multiple modes within the spectral width.

- However, we can also get this result semiclassically by letting the driving field $E(t)$ be an incoherent superposition of a range of frequencies $\delta\omega_0$. After driving for a time t, we can resolve frequency differences of $1/t$, and hence the number of resolvable frequency bins is $\omega_0 t$. These average with random phases, giving a suppression of $1/\sqrt{\omega_0 t}$ to the rms amplitude. Then the expected probability only grows as t. In the long time limit, the probabilities of occupying the excited state eventually saturates at $1/2$.

- This is again a bit complicated to think about, so below we will ignore this and treat the driving as perfectly monochromatic, which is perfectly correct as long as we drive for a time $t \ll 1/\delta\omega_0$. Now, for strong monochromatic fields close to resonance, the transition probability does not remain small for all times, so first-order perturbation theory breaks down. Rabi found that one could derive a more exact result using only the rotating wave approximation, getting a qualitatively similar conclusion.

- To do this, we return to the time-dependent Schrodinger equation, and solve it explicitly. For convenience, let's pull out the phases, giving

$$\dot{c}_0 = -\frac{i}{2\hbar}\mathcal{M}e^{i\Delta t}c_1, \quad \dot{c}_1 = -\frac{i}{2\hbar}\mathcal{M}e^{-i\Delta t}c_0$$

  using the rotating wave approximation. Plugging the first equation into the second,

$$\ddot{c}_1 + i\Delta\dot{c}_1 + \frac{1}{4}\frac{\mathcal{M}^2}{\hbar^2}c_1 = 0.$$

- This is a linear differential equations whose characteristic equation has roots

$$c_1(t) = A_+ e^{i\lambda_+ t} + A_- e^{i\lambda_- t}, \quad \lambda_\pm = \frac{1}{2}\left(\Delta \pm \sqrt{\Delta^2 + \mathcal{M}^2/\hbar^2}\right).$$

  Applying the initial conditions, we find

$$c_1(t) = \frac{i\mathcal{M}}{\Omega\hbar}e^{i\Delta t/2}\sin(\Omega t/2), \quad \Omega_R = \sqrt{\Delta^2 + \mathcal{M}^2/\hbar^2}, \quad P_1(t) = \frac{|\mathcal{M}|^2}{\hbar^2}\frac{\sin^2(\Omega t/2)}{\Omega^2}$$

  where Ω_R is the Rabi frequency.

- We see that for large detuning Δ, the result matches first-order perturbation theory, while otherwise it differs in amplitude and frequency, but not qualitatively. For example, for exact resonance, $\Delta = 0$, the amplitude of oscillation is unity, and the Rabi frequency is $\Omega_R = \mathcal{M}/\hbar$.

- This behavior is called Rabi flopping, or Rabi oscillations. If we now account semiclassically for the finite spread in frequencies, we find as in first-order perturbation theory that the Rabi oscillations eventually "collapse", leaving a probability of $1/2$ for each state.

- For transitions in the visible-frequency range, observations of Rabi flopping require powerful laser beams, as otherwise they damp out before they can build up.

- In NMR experiments, applying a perturbation on resonance for time π/Ω_R, which flips the populations, is called a π pulse. Applying a perturbation on resonance for time $\pi/2\Omega_R$ turns the ground state into an equal position of the ground and excited states, and is called a $\pi/2$ pulse.

Now we discuss the same driven two-level system in the context of NMR. This will be completely mathematically identical, but lead to a different and useful visualization.

- First, consider a classical magnetic moment in a magnetic field. We have

$$\boldsymbol{\tau} = \boldsymbol{\mu} \times \mathbf{B}, \quad \boldsymbol{\mu} = \gamma \mathbf{L}$$

  where γ is called the gyromagnetic ratio. For any magnetic moment built out of classical charges, or more precisely for any classical nonrelativistic system whose charge distribution and mass distribution are proportional, $\gamma = q/2m$.

- Combining these equations, we have

$$\dot{\mathbf{L}} = -\gamma \mathbf{B} \times \mathbf{L}$$

  which is spin precession at the Larmor frequency

$$\Omega_L = \gamma B.$$

- Quantum mechanically e.g. for an electron, the magnetic moment is

$$\mu = \frac{e}{2m} g \frac{\hbar}{2} \approx \frac{e\hbar}{2m} \equiv \mu_B$$

  where $g \approx 2$ is the electron g-factor, a fudge factor that corrects the classical result, and $\hbar/2$ is its angular momentum. Hence the electron has gyromagnetic ratio

$$\gamma \equiv \frac{qg}{2m} \approx \frac{q}{m}.$$

- Experimentally, we find that $g_p = 5.6$ for the proton and $g_n = -3.7$ for the neutron. Numerically, the gyromagnetic ratios are

$$\gamma_e/2\pi = 2.8\,\text{MHz/gauss}, \quad \gamma_p/2\pi = 4.2\,\text{kHz/gauss}$$

  where we have divided by 2π to convert to ordinary frequency.

- The Larmor frequency is on the order of the cyclotron frequency for free charges,

$$\Omega_c = \frac{qB}{m}.$$

  Since the electron has $g \approx 2$ and spin $1/2$, this is about equal to the Larmor frequency, which might be helpful for intuition.

Note. A common heuristic argument for $g = 2$. This fact implies that the electron precesses twice as fast as the classical expectation. To understand this, note that for a classical spin, we are measuring its rotation with respect to a non-rotating spin, say one with no magnetic moment. However, for a spin $1/2$ particle, we have to compare the phases of the spin up and spin down states, which evolve with opposite phase shifts. So the relative phase precesses twice as fast as expected.

This is an elegant argument, but it has the minor disadvantage of being wrong. For example, it implies all spin $1/2$ particles would have $g = 2$, but the proton and neutron don't.

Now we analyze our classical spinning magnetic moment in a rotating frame.

- If some vector $\mathbf{A}$ rotates with angular velocity $\mathbf{\Omega}$, then

$$\frac{d\mathbf{A}}{dt} = \mathbf{\Omega} \times \mathbf{A}.$$

  Therefore, in general, for any vector $\mathbf{A}$, we have

$$\left.\frac{d\mathbf{A}}{dt}\right|_{\text{in}} = \left.\frac{d\mathbf{A}}{dt}\right|_{\text{rot}} + \mathbf{\Omega} \times \mathbf{A}$$

  where the left-hand side is the rate of change of $\mathbf{A}$ in the inertial lab frame, and the right-hand side contains the rate of change of $\mathbf{A}$ in a frame rotating with angular velocity $\mathbf{\Omega}$.

- More generally, this yields the operator equation

$$\left.\frac{d}{dt}\right|_{\text{rot}} = \left.\frac{d}{dt}\right|_{\text{in}} - \mathbf{\Omega} \times .$$

  In the rotating frame, our evolution equation for the spin becomes

$$\left.\frac{d\mathbf{L}}{dt}\right|_{\text{rotating}} = \gamma \mathbf{L} \times (\mathbf{B} + \mathbf{\Omega}/\gamma) = \gamma \mathbf{L} \times \mathbf{B}_{\text{eff}}.$$

  Therefore, the spin is static if $\mathbf{\Omega} = -\gamma \mathbf{B}$, which implies that in the lab frame, it precesses with $\mathbf{\Omega} = -\gamma \mathbf{B}$, just as we found earlier.

- Now consider a magnetic moment in a magnetic field $\mathbf{B} = B\hat{\mathbf{z}}$, precessing at the Larmor frequency, which we call ω. Suppose we superpose an additional rotating field of the form

$$\mathbf{B}_1(t) = B_1(\cos \omega t\, \hat{\mathbf{x}} - \sin \omega t\, \hat{\mathbf{y}})$$

  which also rotates at the Larmor frequency.

- In the frame rotating at the Larmor frequency, the original static field disappears, and the new field is static,

$$\mathbf{B}_{\text{rot}} = B_1\, \hat{\mathbf{x}}'.$$

  Therefore the spin flips over with the on-resonance Rabi frequency

$$\Omega_{R,0} = \gamma B_1$$

  Even though the applied field can be very small, it still has a large effect, because it resonates with the Larmor precession. We can flip the spin with a pulse of time $\pi/\Omega_{R,0}$.

- All of these results remain true in quantum mechanics. The equations of motion we started with simply coincide with the Heisenberg equations. Going to the rotating frame is simply working in interaction picture. The difference is that we now have a way to visualize any quantum two-state system in three dimensions.

- In practice, we usually use a linearly oscillating field at the resonant frequency? This is a superposition of a resonantly rotating field, and a field rotating the opposite way. Transforming to the rotating frame, we get the same result as before, plus an extra field rotating at $2\omega_0$, and ignoring this rapidly rotating field is again the rotating wave approximation.

- Our geometrical picture now gives an easy derivation of the Rabi frequency off resonance. Suppose our driving field has frequency ω. In the frame rotating with that driving field, we have the static field

$$\mathbf{B}_{\text{rot}} = B_1\hat{\mathbf{x}}' + (B_0 - \omega_0/\gamma)\hat{\mathbf{z}}'.$$

  The spin precesses about this field, at the Rabi frequency

$$\Omega_R = \sqrt{(\omega - \omega_0)^2 + (\gamma B_1)^2}$$

  which precisely matches our earlier result, in different conventions. One can do a bit more geometry to find the maximum amplitude of oscillations of μ_z as well.

The intuition above allows us to understand another way of flipping spins, rapid adiabatic passage.

- In rapid adiabatic passage, a spin can be flipped by slowly scanning the driving frequency ω_0 past the Larmor frequency ω. It turns out that this will cause the spin to flip. This technique particularly useful if we don't precisely know the Larmor frequency.
- To see why, suppose the spin begins parallel to the static field $\mathbf{B}_0$, which is indeed what happens after thermalization. Now, if we turn on the rotating field $\mathbf{B}_1$ well below resonance, then in the frame rotating with $\mathbf{B}_1$, the spin precesses in a tight circle about an effective field, which is almost parallel to the original $\mathbf{B}_0$.
- As we slowly sweep the frequency up past the resonance, the effective field goes from pointing up, to pointing horizontally (at resonance), to pointing down. During the entire time, the spin precesses tightly about the effective field. Therefore, the magnetic moment ends up flipped!
- Many variants of this procedure also work. We can sweep the magnitude of the static field, passing the Larmor frequency past the rotating field's frequency. We can also sweep the frequency from high to low; the logic is identical, except the moment rotates in a tight circle *against* the effective field.
- The sweeping must be "rapid", in the sense that it is faster than relaxation/decoherence processes. It must also be adiabatic, to keep the spin following the field. This occurs if the fractional change in the field during one period is small,

$$\frac{d\omega_0}{dt} \ll \Omega_R^2.$$

  This is most restrictive when the driving field is on resonance, giving

$$\frac{d\omega_0}{dt} \ll \Omega_{R,0}^2 = (\gamma B_1)^2.$$

- Understanding the error in the adiabatic theorem is best done quantum mechanically. In the interaction picture, the situation above corresponds to the time-dependent Hamiltonian

$$H = \frac{\hbar}{2}\begin{pmatrix} \delta(t) & \Omega_{R,0} \\ \Omega_{R,0} & -\delta(t) \end{pmatrix}$$

  where $\delta(t)$ is varied from $+\infty$ to $-\infty$. The two energy levels have an avoided crossing, separated by $\hbar\Omega_{R,0}$ at the point $\delta = 0$. If δ is varied slowly enough that the adiabatic theorem holds, then the spin state ends up flipped.

- The Landau–Zener problem asks for the probability of jumping through the avoided crossing, i.e. quantifying the error in the adiabatic theorem. For uniform $d\delta/dt$, it turns out that

$$P = \exp\left(-\frac{\pi}{2}\frac{\Omega_{R,0}^2}{d\delta/dt}\right).$$

  As a check, note that $P = 1$ for fast flipping. This makes sense, because the spin has no time to react, and it stays in the same direction; hence it ends up in the opposite eigenstate.

4.2 The Jaynes–Cummings Model

The Jaynes–Cummings model is the fully quantum mechanical analogue of the Rabi model.

- For an atom in free space, there is a continuum of photon modes, so it is difficult to treat the electromagnetic field quantum mechanically. However, for an atom in a cavity, one can engineer only one mode of the cavity to be on resonance.

- For clarity, we rename the atomic states $|g\rangle$ and $|e\rangle$. The interaction is

$$H_{\text{int}} = -\mathbf{d}\cdot\mathbf{E} \propto (|g\rangle\langle e| + |e\rangle\langle g|)(a + a^\dagger)$$

  where the constant of proportionality involves the electric field evaluated at the atom's position, and there are no $|g\rangle\langle g|$ or $|e\rangle\langle e|$ terms because photon emission flips the parity.

- It is convenient to introduce the atomic transition operators

$$\sigma_+ = |e\rangle\langle g|, \quad \sigma_- = |g\rangle\langle e|$$

  and the inversion operator

$$\sigma_3 = |e\rangle\langle e| - |g\rangle\langle g|.$$

  These obey the commutation relations

$$[\sigma_+, \sigma_-] = \sigma_3, \quad [\sigma_3, \sigma_\pm] = \pm 2\sigma_\pm.$$

- The interaction Hamiltonian takes the form

$$H_{\text{int}} = \hbar\lambda(\sigma_+ + \sigma_-)(a + a^\dagger).$$

  We shift the energy to be zero midway between the two atomic states when there are no photons present, so

$$H_0 = \frac{1}{2}\hbar\omega_0\sigma_3 + \hbar\omega a^\dagger a.$$

- Now suppose the interaction is weak. Then the operators $\sigma_\pm$ and a and $a^\dagger$ evolve with the usual phases,

$$a(t) = a(0)e^{-i\omega t}, \quad \sigma_\pm(t) = \sigma_\pm(0)e^{\pm i\omega_0 t}.$$

  Assuming we are near resonance, $\omega \approx \omega_0$, two of the terms in the interaction Hamiltonian oscillate slowly and two oscillate quickly. Dropping the latter is again the rotating wave approximation, and gives

$$H = H_0 + \hbar\lambda(\sigma_+ a + \sigma_- a^\dagger).$$

  This is the Jaynes–Cummings model.

- Now, we note that the electron number and excitation number

$$P_E = |e\rangle\langle e| + |g\rangle\langle g|, \quad N_e = a^\dagger a + |e\rangle\langle e|$$

are both conserved. The Hamiltonian can be written as

$$H = \left(\hbar\omega N_e + \hbar\left(\frac{\omega_0}{2} - \omega\right)P_E\right) - \hbar\Delta\sigma_3 + \hbar\lambda(\sigma_+ a + \sigma_- a^\dagger).$$

We can shift the zero of energy again to remove the first term.

We now consider some simple examples on resonance, $\Delta = 0$.

- Let the initial state be $|i\rangle = |e\rangle|n\rangle$. Note that the Hamiltonian only mixes this state with $|f\rangle = |g\rangle|n+1\rangle$. For the free Hamiltonian, these states have precisely the same energy.
- It is then straightforward to solve for the time evolution exactly. Let

$$|\psi(t)\rangle = c_i(t)|i\rangle + c_f(t)|f\rangle.$$

The amplitude rotates between these states with angular frequency $\lambda\sqrt{n+1}$,

$$c_i(t) = \cos(\lambda t\sqrt{n+1}), \quad c_f(t) = -i\sin(\lambda t\sqrt{n+1}).$$

- We can define a quantum Rabi frequency

$$\Omega(n) = 2\lambda\sqrt{n+1}$$

which of course limits to the expected result $\mathcal{M}/\hbar$ in the limit of high n. One difference is that the Rabi oscillations occur even when $n = 0$, where semiclassically there would be no field at all, as in this case they are started via spontaneous emission.

- A more surprising result occurs if we do not start with a number eigenstate. Let us take the initial state

$$|\psi(0)\rangle = \left(\sum_{n=0}^{\infty} c_n|e\rangle|n\rangle\right).$$

Then by similar reasoning, the state at later times is

$$|\psi(t)\rangle = |\psi_g(t)\rangle|g\rangle + |\psi_e(t)\rangle|e\rangle$$

where

$$|\psi_g(t)\rangle = -i\sum_{n=0}^{\infty} c_n \sin(\lambda t\sqrt{n+1})|n+1\rangle, \quad |\psi_e(t)\rangle = \sum_{n=0}^{\infty} c_n \cos(\lambda t\sqrt{n+1})|n\rangle.$$

- One useful quantity is the atomic inversion,

$$W(t) = \langle\psi(t)|\sigma_3|\psi(t)\rangle = \sum_{n=0}^{\infty} |c_n|^2 \cos(2\lambda t\sqrt{n+1}).$$

For concreteness, we take a coherent state with parameter α, where $\overline{n} = |\alpha|^2$, giving

$$W(t) = e^{-\overline{n}} \sum_{n=0}^{\infty} \frac{\overline{n}^n}{n!} \cos(2\lambda t\sqrt{n+1}).$$

- Now consider how the amplitude changes over time. For high α, the coefficients are distributed around $\overline{n}$ as a Gaussian with width $\sqrt{\overline{n}}$. Therefore, the fractional spread in the frequencies is

$$\frac{\sqrt{n+\sqrt{n}}}{\sqrt{n}} - 1 \sim \frac{1}{\sqrt{n}}.$$

  Therefore, after $\sqrt{n}$ oscillation cycles, the oscillations damp down since the components get out of phase. This corresponds to a collapse time $t_c \sim 1/\lambda$.

- This can be understood semiclassically. If we drive with light with bandwidth $\delta\omega$, damping occurs for $t \gtrsim 1/\delta\omega$. The logic here is precisely the same, except that the frequency spread is due to the inherent amplitude uncertainty of light. But we can ultimately ascribe this to the Heisenberg uncertainty principle for whatever matter system created the light.

- However, there is a further effect that directly reveals the quantized nature of the electromagnetic field: long after the collapse, the oscillations experience a "revival". This is because for high $\overline{n}$, the frequencies are approximately evenly spaced, by applying a tangent line approximation to $\sqrt{n}$. The frequency differences are determined by

$$\sqrt{n+1} - \sqrt{n} \sim \frac{1}{\sqrt{n}}.$$

  Then the oscillations get back into phase at time

$$t_r \sim \frac{\sqrt{\overline{n}}}{\lambda}.$$

  The collapses and revivals continue, until eventually the error in the tangent line approximation damps them out.

- Collapses and revivals also occur, to a lesser extent, in variations of the Jaynes–Cummings model with more field modes. In the limit of infinitely many, we of course recover the semiclassical exponential decay.

Next, we introduce the idea of dressed states.

- Dressed state are the eigenstates of the Jaynes–Cummings model. We return to the basic Hamiltonian, now allowing a finite detuning,

$$H = \frac{1}{2}\hbar\omega_0\sigma_3 + \hbar\omega a^\dagger a + \hbar\lambda(a\sigma_+ + a^\dagger\sigma_-).$$

  The interaction terms only mix pairs of states, such as $|e\rangle|n\rangle$ with $|g\rangle|n+1\rangle$. States like these are called the "bare" states, since they are the eigenstates when the interaction is off.

- Restricting to the subspace spanned by

$$|\psi_{1n}\rangle \equiv |e\rangle|n\rangle, \quad |\psi_{2n}\rangle \equiv |g\rangle|n+1\rangle$$

  the Hamiltonian is a 2×2 matrix,

$$H^{(n)} = \hbar \begin{pmatrix} n\omega + \frac{1}{2}\omega_0 & \lambda\sqrt{n+1} \\ \lambda\sqrt{n+1} & (n+1)\omega - \frac{1}{2}\omega_0 \end{pmatrix}.$$

- The energy eigenvalues are

$$E_{\pm}(n) = \left(n + \frac{1}{2}\right)\hbar\omega \pm \frac{1}{2}\hbar\Omega_n(\Delta), \quad \Omega_n(\Delta) = \sqrt{\Delta^2 + 4\lambda^2(n+1)}$$

  where Ω_n is the quantum analogue of the Rabi frequency. The corresponding eigenstates are called "dressed" states. The splitting of the dressed states can be thought of as a kind of "AC" or "dynamic" Stark effect. In the case $\Delta = 0$, the bare states are initially degenerate, so this provides an example of how interactions generically break degeneracies.

4.3 Atoms in Cavities

Before beginning, it is useful to review a simple optical cavity.

- Consider a planar cavity of length L consisting of two parallel mirrors. A phase shift of either zero or π occurs upon reflection with each mirror, but since reflections come in pairs, this won't matter. Hence the only property of the mirrors that matter is their reflectivity R_i, i.e. the proportion of energy reflected by each mirror individually.

- By summing the infinite series of possible paths through the mirror, one finds a transmission

$$T = \frac{1}{1 + (4\mathcal{F}^2/\pi^2)\sin^2(\phi/2)}$$

  where ϕ is the round trip phase shift and $\mathcal{F}$ is the finesse,

$$\phi = \frac{4\pi L}{\lambda}, \quad \mathcal{F} = \frac{\pi(R_1R_2)^{1/4}}{1 - \sqrt{R_1R_2}}.$$

  If the reflectivities are equal and high, which we assume from here on, then $\mathcal{F} \approx \pi/(1-R)$.

- We see that resonance occurs whenever

$$L = \frac{m\lambda}{2}$$

  which corresponds to a round trip phase shift of $2\pi m$. In this case, there is perfect transmission through the cavity. The full width at half maximum of this resonance is inversely proportional to $\mathcal{F}$, so high-finesse cavities can be used for spectroscopy.

- Now consider this optical cavity by itself, with a light source inside. The resonant modes of the cavity have frequency

$$\omega_m = \frac{m\pi c}{L}.$$

  Photons bounce inside the cavity on the order of $1/(1-R) \sim \mathcal{F}$ times. Thus, on resonance, the light intensity is enhanced by $\mathcal{F}$, so the height of a resonance peak is inversely related to its width. For light far off resonance, the intensity experiences destructive interference, giving a penalty of order $1 - R$.

We now briefly discuss atoms in cavities.

- When an atom is placed in a cavity and an atomic transition is on resonance with a cavity mode, then the system is described by three parameters: the photon decay rate of the cavity κ, the non-resonant decay rate γ, and the atom-photon coupling parameter g_0, which has the dimensions of a rate.

- The photon decay rate κ is determined by the width of the resonance,

$$\Delta\omega = \frac{\omega}{Q}$$

  where the quality factor generalizes the finesse above. The non-resonant decay rate is due to the atom emitting into other modes, and decaying or becoming excited to other levels. (For example, many of these cavities are just two mirrors, so photons can escape if the atom emits them "sideways".)

- We are in the strong coupling regime when $g_0 \gg \max(\kappa, \gamma)$, and the weak coupling regime when the opposite is true. In the strong coupling regime, photons emitted by the atom can be reabsorbed by it, leading to coherent Rabi oscillations. It is easier to reach strong coupling with $N \gg 1$ atoms in the cavity, because the decay rates only grow as N, while the photon/atom coupling rate grows as N^2.

- At weak coupling, a photon emitted by the atom effectively never returns, so the most striking effect is the Purcell effect: the cavity can modify the density of states, changing the rate of spontaneous emission from the atom. In this case, we can treat the atom perturbatively.

- We recall the spontaneous emission rate in free space. By Fermi's golden rule, we have

$$\Gamma \propto |M|^2 g(\omega).$$

  This is an indeterminate form, which is regulated by placing the atom in a large box of volume V. Then $|M|^2 \propto 1/V$, while

$$g(\omega) = \frac{\omega^2 V}{\pi^2 c^3}.$$

  The unphysical volume V cancels out of the result.

- Now consider an atom in a cavity of volume V_0, so that the lossless version of the cavity has exactly one relevant mode, at frequency ω_c. Handling the spontaneous emission rate in perturbation theory is tricky, because there's only one relevant cavity mode rather than a continuum, and moreover this mode decays to the environment, giving it a finite width $\Delta\omega_c$.

- However, a detailed analysis shows that these two subtleties cancel each other out: it turns out we can compute Γ by applying Fermi's golden rule, and taking the density of states to be a Lorentzian with unit area and full width at half maximum $\Delta\omega_c$,

$$g(\omega) = \frac{2}{\pi} \frac{\Delta\omega_c}{4(\omega - \omega_c)^2 + (\Delta\omega_c)^2}.$$

- Since $|M|^2 \propto 1/V_0$, the spontaneous emission rate is enhanced on resonance by the Purcell factor,

$$F = \frac{Q}{4\pi^2} \frac{\xi}{1/3} \frac{\lambda^3}{V_0}$$

where we defined $\lambda = 2\pi c/\omega$. The factor $\xi \leq 1$ accounts for the alignment of the atom's dipole moment with the mode's local field, with an effective factor of $1/3$ for free space because of the three dimensions of space.

- The Purcell effect was first observed in the early 1980s in superconducting cavities, then in the late 1980s in the optical regime. In the 1990s, it was also observed in semiconductor heterostructures, which were used to create microcavities. One possible application of the Purcell effect is the creation of low threshold lasers; for instance, the optical density of states can be precisely controlled in a photonic crystal.

- As noted above, the main non-semiclassical feature of the strong coupling regime, described by the Jaynes–Cummings model, is the presence of vacuum Rabi splitting. This was first demonstrated in the 1990s. A probe beam was sent through an optical cavity; as shown above, transmission occurs near the resonant frequencies of the cavity. When atoms are introduced into the cavity and the experiment is repeated, the transmission peaks are split in half.

- Such an effect is easier to see when there are many atoms present. The first observation of vacuum Rabi splitting using only single atoms was in the mid 2000s.

5 Open Systems

5.1 Decoherence

In this section, we study the general phenomenon of decoherence.

- Decoherence explains how a microscopically quantum world is compatible with a macroscopically classical world. Realistic, macroscopic systems quickly become irreversibly entangled with their environments. In classical mechanics, interaction with the environment could simply be parametrized as a perturbation of the system's state, but in quantum mechanics the situation is far more severe: a system entangled with its environment cannot be described as having a (pure) state of its own at all, even if it is subsequently removed from the environment. (This is related to the so-called "nonlocality" of quantum correlations.) This washes out quantum interference effects in the system.

- In more detail, a full theory of decoherence would also provide "environment-induced superselection (einselection) rules", i.e. criteria for which system states, called pointer states, are stable under interaction with the environment. An important point here is that the superselection rules are "purely" quantum mechanical, in the sense that the classical world emerges from only quantum rules, without any classical/quantum boundary put in by hand.

- For example, collisional decoherence can occur when external particles collide with the system. The resulting decoherence effects can be large even if the naive classical effects are very weak: a light external particle colliding with a heavy one can transmit negligible momentum, yet still effectively collapse a superposition of positions. This process is so effective that it occurs essentially instantly for macroscopic objects, even accounting for only CMB radiation.

- Sometimes, it's claimed that decoherence "solves" the interpretation of quantum mechanics. It's true that it removes much of the mystique behind "collapse", in the sense that collapse on measurement can be effectively derived from the rules of quantum mechanics, without any additional mechanisms; we no longer need to postulate inherently "classical" measurement apparatuses. It also tells us why we don't expect to see "Schrodinger's cat" states in nature.

- However, many issues remain. If one does not have a hard classical/quantum boundary, we potentially have to consider ourselves as quantum systems, introducing the question of why nobody perceives *themselves* to be in a superposition after performing a measurement, but instead sees apparently probabilistic but definite outcomes. This "problem of definite outcomes" leads to paradoxes, exemplified in "Wigner's friend" thought experiments. Some interpretations attempt to resolve it by adding additional, "objective" collapse mechanisms.

- Furthermore, decoherence does not resolve the ontological problems of interpretation. This is the problem of *what* a quantum state describes, e.g. whether it describes something "physical" or simply an observer's knowledge of the system's state, and whether it contains "all" of the physical information that exists about a system.

Example. Suppose a system is prepared in a superposition of two orthogonal states, $(|a_1\rangle+|a_2\rangle)/\sqrt{2}$. The system interacts with the environment, so that the final state is

$$\frac{|a_1 b_1\rangle + |a_2 b_2\rangle}{\sqrt{2}}.$$

The results of all measurements that can be performed on the system alone can be encapsulated with the reduced density operator, which is

$$\rho_A = \frac{1}{2}\left(|a_1\rangle\langle a_1| + |a_2\rangle\langle a_2| + |a_1\rangle\langle a_2|\langle b_2|b_1\rangle + |a_2\rangle\langle a_1|\langle b_1|b_2\rangle\right)$$

or in matrix form in the relevant basis,

$$\rho_A = \frac{1}{2}\begin{pmatrix} 1 & \langle b_1|b_2\rangle \\ \langle b_2|b_1\rangle & 1 \end{pmatrix}.$$

Because the off-diagonal elements of the density matrix in this basis quantity the amount of interference that can be measured between the states $|a_1\rangle$ and $|a_2\rangle$, interference effects are smaller the better the states $|b_1\rangle$ and $|b_2\rangle$ can be distinguished.

To be very explicit, consider the case of the double slit experiment, where $|a_1\rangle$ and $|a_2\rangle$ correspond to the particle going through either slit, with corresponding wavefunctions $\psi_i(x)$ at the screen. The intensity at the screen is

$$\langle x|\rho_A|x\rangle = \frac{1}{2}|\psi_1(x)|^2 + \frac{1}{2}|\psi_2(x)|^2 + \mathrm{Re}(\psi_1(x)\psi_2^*(x)\langle b_2|b_1\rangle)$$

so the fringe contrast depends directly on $\langle b_2|b_1\rangle$. For example, here the $|b_i\rangle$ could be states of the double slit apparatus itself. If the particle delivers enough momentum to the apparatus to distinguish which slit it went through, then no interference pattern appears; however in reality, the impulses are small compared to the preexisting momentum uncertainty of the apparatus, so $\langle b_2|b_1\rangle$ is high. Alternatively, the $|b_i\rangle$ could represent the states of the surrounding air molecules. If the molecules begin in some product state $\prod_i |e_i\rangle$, then we have

$$|b_1\rangle = \prod_i |e_i'\rangle, \quad |b_2\rangle = \prod_i |e_i''\rangle, \quad \langle b_1|b_2\rangle = \prod_i \langle e_i'|e_i''\rangle$$

where the primed states are the state of each molecule after interacting with a particle going through the top or bottom slit; this is the setup of collisional decoherence. In this case the "which path" information has leaked into the environment, and the decoherence is thermodynamically irreversible.

Intuitively, we can think of the environment as trying to measure whether the system is in state $|a_1\rangle$ or $|a_2\rangle$, with an effective collapse occurring if the environment succeeds; in that case the environment acts as an ideal von Neumann measurement apparatus. This is an idealization of measurement because many realistic measurements are destructive, e.g. most photon detectors destroy the photon.

Note. One has to be careful with reduced density matrices, because they lump together different things. For example, suppose that one flips a coin, then uses it to determine whether to send a particle through the top or bottom slit. This leads to the same density matrix as a particle that is sent through a superposition of both slits, but whose presence is detected at the slits. It is true that these lead to the same pattern at the screen, but the latter is different in principle, because depending on the detection mechanism, we might be able to *erase* the which-path information, restoring the interference. Eliding this distinction leads to paradoxes like the quantum eraser.

In other cases, one can numerically simulate the decay of off-diagonal terms in the density matrix due to decoherence by using stochastic "kicks" to the density matrix. However, this is also a bit misleading because the relevant dynamics are perfectly unitary. Furthermore, significant decoherence can happen even if the "kicks" to the system itself are negligible, because what matters is the overlap of the *environmental* states.

We now make some more concrete remarks about decoherence.

- The general setup is that we start with an unentangled system and environment, with the system in state $|s_i\rangle$, and the environment in the "ready" state. This evolves to the state $|s_i\rangle|E_i(t)\rangle$. In many cases, the overlaps between environmental states decay exponentially,

  $$\langle E_i(t)|E_j(t)\rangle \propto e^{-t/\tau_d}$$

  where τ_d is the decoherence timescale. For multiple sources of decoherence, the overlaps are usually just multiplied.

- In particular, in the case of collisional decoherence, we take the system states to be position eigenstates, and

  $$\langle E_x(t)|E_{x'}(t)\rangle \propto e^{-\Lambda|x-x'|^2 t}.$$

- Note that the set of pointer states is not a vector space. Recall the ideal double slit case,

  $$|\psi_i\rangle|E_0\rangle \to |\psi_i\rangle|E_i\rangle, \quad \langle E_1|E_2\rangle = 0.$$

  This means the superpositions of the $|\psi_i\rangle$ are not stable. But now consider the states

  $$|\psi_\pm\rangle = \frac{1}{\sqrt{2}}(|\psi_1\rangle \pm |\psi_2\rangle).$$

  These evolve as

  $$|\psi_\pm\rangle|E_0\rangle \to \frac{1}{\sqrt{2}}(|\psi_1\rangle|E_1\rangle \pm |\psi_2\rangle|E_2\rangle)$$

  and one cannot distinguish them by just measuring the environment. This means some superpositions of the $|\psi_\pm\rangle$ could be stable, which makes sense, because the $|\psi_i\rangle$ states are.

- More generally, one could have "pointer subspaces" (called decoherence free subspaces in the context of quantum error correction), within which superpositions are stable, and across which superpositions are not stable. Sometimes, as in the double-slit example we just considered, one has a "pointer basis", i.e. a set of linearly independent pointer states that span the Hilbert space. In other cases one can have an overcomplete set of pointer states, such as the coherent states.

- This idea puts the older idea of superselection rules on more solid ground. Superselection rules are simply postulates that certain physically unobserved superpositions, such as a superposition between two particles of different charge, cannot occur. They can now be *derived* through decoherence. In this case, such a superposition would decohere because of the differing Coulomb fields at large distances.

- To determine the set of pointer states, write the total Hamiltonian as

  $$H = H_S + H_E + H_{\text{int}}.$$

  There are a few tractable limits. In the quantum measurement limit, H_{int} dominates, and the pointer states are the ones obeying

  $$\langle s_j|H_{\text{int}}|s_i\rangle = 0$$

  where the right-hand side is the zero operator on the environment Hilbert space.

- In this case, the time evolution is

$$e^{-iH_{\text{int}}t}|s_i\rangle|E_0\rangle = \lambda_i|s_i\rangle e^{-i\langle s_i|H_{\text{int}}|s_i\rangle t}|E_0\rangle \equiv |s_i\rangle|E_i(t)\rangle$$

  which is exactly the condition to be a pointer state.

- Another equivalent way to phrase this is in terms of pointer observables. Given the pointer states $|s_i\rangle$, a corresponding pointer observable commutes with the interaction Hamiltonian,

$$O_S = \sum_i o_i|s_i\rangle\langle s_i|, \quad [O_s, H_{\text{int}}] = 0.$$

- In many simple models, the interaction Hamiltonian has the product form

$$H_{\text{int}} = S \otimes E.$$

  In this case S is a pointer observable; it is the quantity that the environment effectively measures. If S has degenerate eigenvalues, one gets corresponding pointer subspaces. In cases where we have multiple terms,

$$H_{\text{int}} = \sum_\alpha S_\alpha \otimes E_\alpha$$

  simultaneous eigenvectors of the S_α are pointer states.

- Another tractable special case is when H_S is the dominant term, i.e. when the separations between energies of system states are much larger than any other energies in the problem. In this case, the environment can only effectively measure constants of the motion. Thus, in the absence of other conserved quantities, energy eigenstates are generically the pointer states.

- However, in many physical applications, none of the terms in the Hamiltonian are dominant, and we need to resort to numeric methods. For example, one can quantify the robustness of a state using the time evolution of the corresponding reduced density matrix, e.g. the decay of $\operatorname{tr}\rho_S^2$ or the growth of $-\rho_S \log \rho_S$. In these cases, being a "pointer state" is not an all-or-nothing classification; instead there are simply states that are more or less robust against interaction with the environment, with different measures of robustness generally agreeing.

- As one example, consider collisional decoherence. If one treated H_{int} as dominant, the pointer states would be position eigenstates. But this isn't the complete picture, because classically we expect definite momentum as well. Measurement of momentum occurs because the momentum of the particle affects its position by time evolution under H_S.

Note. As we mentioned above, decoherence gets trickier to apply the more one "zooms out". For example, we now have a good idea of how non-pointer system states get entangled with the environment, and hence why pointer states are more stable. But an additional feature of classical physics is that many different observers can measure the same system without disturbing it. Therefore, the "which path" information of a pointer state must be redundantly encoded into the environment for the state to truly act "classically". For example, in collisional decoherence, many environment particles scatter off the system. An observer can infer the approximate position of the system by looking at only some of these scattered particles, while another observer can do the same with others. Understanding how this happens is the task of the "quantum Darwinism" program, and it is technically challenging because the dynamics of the environment are complex and possibly chaotic.

5.2 Models of Decoherence

Example. A simple explicit model for decoherence. The system is a spin 1/2 particle, and the environment is a set of N such spins. The interaction Hamiltonian is

$$H_{\text{int}} = \frac{1}{2}\sigma_z \otimes \sum_i g_i \sigma_z^{(i)}.$$

Assuming this term dominates, the pointer states are the eigenstates of σ_z. Indexing the environmental states by their binary representation $|n\rangle$, an arbitrary initial state can be written as

$$|\psi\rangle = \sum_n c_n|0\rangle|n\rangle + d_n|1\rangle|n\rangle$$

where there are 2^N terms in the sum. The terms in the sum are all eigenstates of H_{int}, so they all evolve with independent phase factors. In particular, if all the c_n and d_n are nonzero, the decoherence rate scales exponentially, as 2^N. At early times, it turns out that for many distributions of the g_i, the coherence falls as $e^{-\Gamma^2 t^2}$. At late times, one can have a Poincare recurrence, but for a general distribution of the g_i this takes time $\tau_r \propto N!$. Another point to note that is that if $H_S \propto \sigma_z$, then $[H_s, H_{\text{int}}] = 0$, so this entire process occurs without any exchange of energy between the system and environment. In other words, decoherence is independent of dissipation.

Note. Throughout, we have assumed that the system and environment begin in a pure state. This entails no loss of generality, because a mixed state can straightforwardly be "purified" by embedding it as a pure state in a larger state space. Explicitly, we can map

$$\rho = \sum_i p_i|\psi_i\rangle\langle\psi_i|$$

to the pure state

$$|\Psi\rangle = \sum_i \sqrt{p_i}|\psi_i\rangle|\phi_i\rangle$$

where the $|\phi_i\rangle$ are orthogonal. This idea was also used in the notes on Quantum Computation. In that context, we also examined tools for analyzing the time evolution of the system's reduced density matrix, such as Kraus operators.

Example. Thermal decoherence. Decoherence can result from the spontaneous emission of photons. For example, consider a superposition of an object of surface area A and temperature T across a distance d. In natural units, the typical emitted photon has energy T, so the rate of photon emission is AT^3. If the photon wavelength $\lambda \sim 1/T$ obeys $\lambda \ll d$, then a single photon emitted can decohere the superposition. However, for $\lambda \gg d$ the decoherence factor per photon is merely $(d/\lambda)^2$. Therefore, the decoherence rate is

$$\Gamma \sim \begin{cases} AT^3 & dT \gg 1, \\ Ad^2T^5 & dT \ll 1. \end{cases}$$

This has been observed for double slit experiments using buckyballs, where T was controlled by an initial heating laser pulse.

Now we consider the important example of collisional decoherence.

- A initially real Gaussian wavepacket of width σ spreads in time,

$$\sigma(t) = \sigma\sqrt{1 + (\hbar t/m\sigma^2)^2}.$$

  Sometimes it is claimed that this doesn't matter because $\hbar$ is tiny, but it actually is disturbingly large. For example, for an electron ($m \approx 10^{-30}\,\text{kg}$) with an initial width of $\sigma = 10^{-10}\,\text{m}$, the wavepacket spreads to the size of the Earth in seconds! This is because the smallness of $\hbar$ is compensated by the smallness of m.

- Because of this, Schrodinger was puzzled over how particles like electrons could ever be treated classically at all. In 1926, he discovered the coherent states of the harmonic oscillator, which do not spread out, and speculated that similar states could be constructed more generally. But as we noted in the notes on Undergraduate Physics, coherent states really are unique to the harmonic oscillator. All wavepackets in a $1/r$ potential spread out, and worse, in chaotic situations the spreading is exponential.

- This puzzle was solved by Joos and Zeh through collisional decoherence, which converts a spread wavepacket into a mixture of wavepackets, with size on the order of the thermal de Broglie wavelength. However, their original calculations were off by numeric factors.

-

5.3 Master Equations

5.4 Quantum Monte Carlo

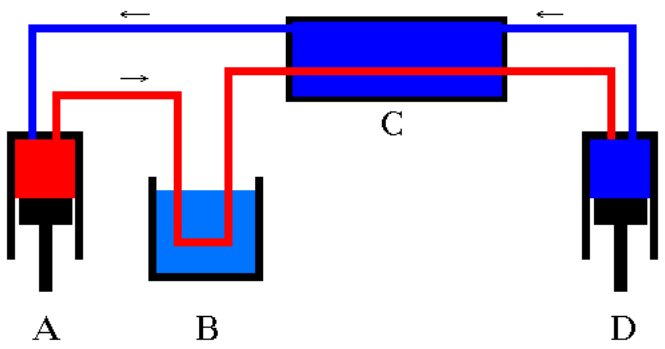
5.5 Quantum Interpretations

6 Precision Atomic Experiments

6.1 Cold Atoms

We begin with a rough overview of cooling methods.

- The theoretically simplest example of a refrigerator is a Carnot cycle operating in reverse. In this case, heat is taken up from a cold reservoir when the gas isothermally expands, further cooling the reservoir.
- This isn't practical for, e.g. a home refrigerator because the heat capacity of the gas is not very large, given the small absolute temperature changes. Instead, home refrigerators typically use a cycle that passes through a liquid-vapor phase transition, which has a large latent heat. Heat is taken from the cold reservoir during the evaporation step. In a compression refrigerator, the vapor is then compressed (at which point work is needed), and heat is dumped to the hot reservoir during the condensation step.
- In the late 19^{th} century, scientists tried to push to significantly lower temperatures. In this case, rather than thinking about a hot and a cold reservoir, we focus on the working fluid itself. The goal is to devise thermodynamic cycles where the working fluid gets colder on each pass through.
- A simple example of such a cycle is the Siemens cycle, shown below.



  First, a gas is compressed by a piston and thereby heated (A). Most of the heat is removed by a cooler (B), and then further cooled by a countercurrent heat exchanger (C). Finally, the gas is again cooled by expansion against a piston at (D). The net effect is that the gas ends up cooler than it started.

- This effect saturates once the temperature of the compressed gas coming out of (A) matches the temperature of the cooler, because at this point, entropy can no longer leave the system. Thus, lower temperatures can be attained by using higher compression pressures. Even lower temperatures can be reached by using the resulting cold fluid as the cooler of a second stage Siemens cycle, which requires the first stage to be substantially larger than the second. Thus, early refrigeration experiments tended to be large and rather dangerous.
- Another important effect is that most real gases cool down when undergoing free expansion because of the attractive intermolecular forces; this is the Joule–Thomson effect. The analogue of the Siemens cycle with the expansion piston replaced with a Joule–Thomson orifice is called the Hampson–Linde cycle, and it is increasingly useful at lower temperatures.

- By chaining together such cycles, many gases were liquefied in the late 19^{th} century, culminating in the liquefaction of helium by Onnes in 1908, and resulting in the discovery of superfluidity and superconductivity. As such, the original applications of low temperature physics are in condensed matter.

- Throughout the 20^{th} century, these techniques were refined to reach even lower temperatures. A dilution fridge begins by liquefying helium, and then reaches even lower temperatures by using the heat of mixing of ^{3}He and ^{4}He. Such fridges reach mK temperatures and are commonly used to cool, e.g. quantum computers and precision physics experiments today.

- Nuclear spins can also be cooled using adiabatic demagnetization; by replacing (H, M) with (P, V), this is perfectly analogous to how an ideal gas is cooled by adiabatic expansion. Starting with nuclear spins cooled by, e.g. liquid helium, adiabatic demagnetization can be used to reach nuclear spin temperatures in the μK range.

In the 1980s, the field of atomic physics leapfrogged past these results, achieving much colder temperatures in gases, as explained below.

- We treat the problem one-dimensionally. A gas at temperature T has

$$\sqrt{\langle v_x^2 \rangle} = \frac{\sqrt{k_B T}}{m}.$$

  We place the atoms in two counterpropagating laser beams, whose frequency is detuned an angular frequency $\Delta < 0$ below an atomic transition. As a result, absorption of a left-moving or right-moving photon is more likely to happen when the atom is moving right or left, respectively. The photon is then emitted in a random direction, resulting in a net damping force, and hence a cooling.

- Quantitatively, let the transition have natural linewidth $\gamma = 1/\tau$. Then the average force due to one of the laser beams is

$$F = -\hbar k R(I, \Delta, v_x)$$

  where k is the photon wavevector, and R is the net absorption rate,

$$R(I, \Delta, v_x) = \frac{\gamma}{2} \frac{I/I_s}{1 + I/I_s + (2(\Delta \pm k v_x)/\gamma)^2}$$

  with the positive sign for the left-moving photons.

- The net absorption rate is the difference of the absorption rate and the stimulated emission rate, because the latter causes anti-damping. The quantity I_s is called the saturation intensity. The expression above can be found by using the standard Lorentzian form of the absorption rate, and accounting for stimulated emission using the Einstein coefficients.

- Adding the effects of the two laser beams and Taylor expanding for small v_x, we find the damping force

$$F_x = -\alpha v_x, \quad \alpha = -\frac{8\hbar k^2 \Delta}{\gamma} \frac{I/I_s}{(1 + I/I_s + (2\Delta/\gamma)^2)^2}.$$

  Because damping occurs for motion in either direction, this setup is known as "optical molasses".

- As the temperature lowers further, to $kv_x \lesssim \gamma$, the net absorption rate becomes roughly equal from both laser beams. The small difference of the net absorption rates still provides a damping, but the many other absorption and emission events cause the momenta of the atoms to perform a random walk, leading to an *increase* of temperature.

- Specifically, if the net absorption rates are equal, then each net absorption and emission event increases $\operatorname{var} p_x = \langle p_x^2 \rangle$ by $2(\hbar k)^2$. Therefore,

$$\frac{dE}{dt} = \frac{1}{2m}\frac{d\langle p_x^2\rangle}{dt} \equiv \frac{D_p}{m} \approx \frac{2\hbar^2 k^2}{m} R(I, \Delta, 0)$$

  where D_p is the momentum diffusion coefficient.

- On the other hand, the damping term reduces the energy at a rate

$$\frac{dE}{dt} = F_x v_x = -\alpha v_x^2.$$

  The minimum attainable temperature, defined by $k_B T/2 = m v_x^2/2$, is hence

$$T = \frac{D_p}{\alpha k_B} = -\frac{\hbar\gamma}{8k_B}\frac{1 + I/I_s + 4\Delta^2/\gamma^2}{\Delta/\gamma}.$$

- This is minimized when $I \ll I_s$ and $\Delta = -\gamma/2$, giving

$$T = \frac{\hbar\gamma}{2k_B}.$$

  This is known as the Doppler cooling limit. In practice, one reaches this limit using "chirp" cooling, beginning with a larger detuning and intensity to speed up the initial cooling, then gradually approaching these parameters.

- It turns out that one can reach even lower temperatures using "Sisyphus cooling". The idea is that the counterpropagating laser beams form a standing wave pattern, which causes a position-dependent energy for the atomic states by the AC Stark effect. The laser can then be tuned so that in the course of one absorption/emission cycle, the atom loses an extra energy equal to the height of the potential hills.

- The fundamental limit on laser cooling is set by the recoil energy due to the absorption and emission of one photon. We have

$$\frac{1}{2}k_B T = \frac{\hbar^2 k^2}{2m}$$

  which gives a recoil temperature of

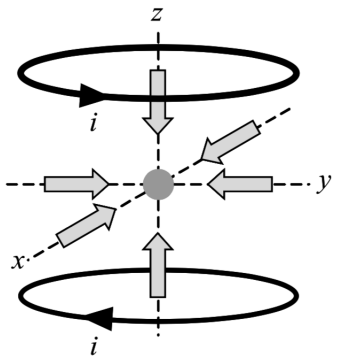
$$T = \frac{h^2}{m k_B \lambda^2}.$$

  A typical Doppler cooling limit is $100\,\mu$K, while a typical recoil limit is 100 nK.

- In principle, we don't need to worry about the other two dimensions: during atomic collisions, the components $\langle p_i^2 \rangle$ will all even out. However, it's common to use a setup with six laser beams, directed along $\pm\hat{\mathbf{x}}$, $\pm\hat{\mathbf{y}}$, and $\pm\hat{\mathbf{z}}$, so that the damping is isotropic.

We now discuss the magnetic traps used to hold the atoms during this process.

- Here we focus on the trapping of neutral atoms. Ion traps are very different, and discussed below. As discussed in the notes on Undergraduate Physics, the primary effect on a neutral atom is through its magnetic moment aligning with or against the magnetic field, and it is possible to produce local minima of B, but not local maxima.
- The energy of a magnetic sublevel in a trap field B is given by the weak-field Zeeman effect, $E = g_J \mu_B B M_J$. Thus, states with $M_J > 0$ can be trapped near a local minimum of B.
- Thus, one possible setup for laser cooling is as shown below.



  The two current loops create a magnetic quadrupole field near the origin, where the magnetic field vanishes.
- This is the simplest option, but it has the disadvantage that the M_J levels are all degenerate at the center of the trap, making it easy to scatter between them. Atoms scattered to negative M_J are repelled away and lost. More sophisticated variants that avoid this problem include the time-averaged orbiting potential trap, and the Ioffe–Pritchard trap.
- Note that Doppler cooling works on sparse atomic gases, while most of the techniques above apply to fluids or solids where interactions are important. This allows exotic low-temperature effects to be studied with more control. For example, superfluid helium involves macroscopically many atoms condensed in the same state, but it's not an ideal Bose–Einstein condensate because interactions are necessary to explain superfluidity at all.
- Much lower temperatures, on the order of 10 nK, are required to make a sparse atomic gas undergo Bose–Einstein condensation, since the de Broglie wavelength needs to be on the order of the atomic separation, but the result is nearly ideal. As a result, this work was given the 1997 and 2001 Nobel prizes.
- Note that there is a gap between the temperature needed for Bose–Einstein condensation and the temperatures that can be reached by laser cooling. The final stage was performed by evaporative cooling, where the depth of the trap is reduced. This allows the most energetic atoms to escape, leaving behind less energetic ones.

Note. Laser cooling and evaporative cooling sometimes trigger suspicion that the Second Law of Thermodynamics is being violated, because in both cases the atom gas ends up with less entropy. In the case of laser cooling, this is because the emitted photons have an increased entropy, due to their random direction. For evaporative cooling, the hotter atoms that escape have an increased entropy because they can now be anywhere on the laboratory floor, rather than in the trap.

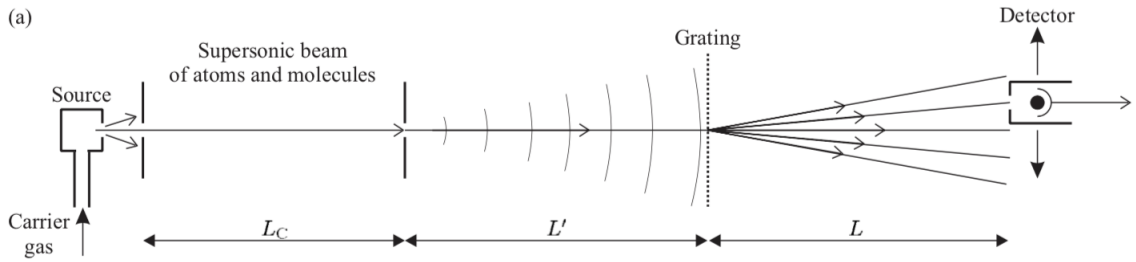
6.2 Atom Interferometry

We begin by introducing the main ideas of atom interferometry.

- Atom interferometry uses the wave-like nature of atoms to perform experiments analogous to light interferometry. These have led to very precise measurements because of the incredibly small wavelengths of atoms, as well as qualitatively new tests of quantum mechanics.
- For a sodium atom with $v = 1000\,\mathrm{m/s}$,

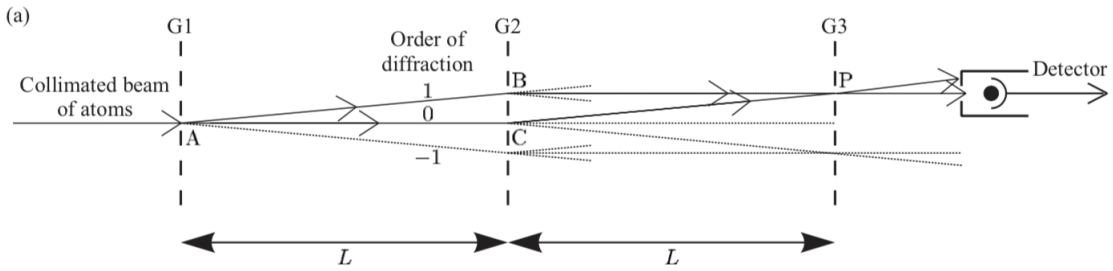
$$\lambda_{\mathrm{dB}} = \frac{h}{mv} = 2 \times 10^{-11}\,\mathrm{m}$$

  which is comparable to that of X-ray radiation.
- X-rays can diffract from crystals, and accordingly neutrons and electrons of similar wavelengths also can be measured to diffract from crystals; this is now a useful measurement technique in condensed matter physics. Scattering matter waves from *engineered* structures took much longer to realize, as this required nanofabrication techniques.
- A schematic of a simple atom interferometry setup is shown below.



  One begins with a heated oven containing sodium metal. A carrier gas of krypton flows through it, carrying along sodium atoms. (One could also just use the oven itself as a source, but the thermally emitted atoms would have a larger spread in velocities.) Between the grating and the source is an intermediate slit, just as in Young's double slit experiment, whose purpose is to increase the spatial coherence of the waves at the grating. In this particular experiment, one can see two families of diffraction maxima, corresponding to Na and Na_2.
- Experiments have also been performed with systems larger than atoms; for example, around the year 2000 interference was observed with C_{60} molecules, also called buckyballs.
- One can summarize the qualitatively new features of atom interferometry as follows: atoms can have short wavelengths and can move slowly compared to the speed of light, making desired phase shifts larger. But unlike light, matter waves disperse in vacuum.
- Newer atom interferometers use laser-cooled atomic gases, which allow for much higher numbers than thermal sources. (If the gases are cool enough to condense into a BEC, then the velocity dispersion is extremely low, and the result is the atomic equivalent of a laser, but this has the disadvantage of greater interactions between the atoms.) The atoms can be detected at the end with laser fluorescence.

Example. Interferometers can be used to measure rotation. This was originally used to test various ether theories, and today is used for engineering applications such as GPS. To see how this can be done with atom interferometry, consider the following three-grating interferometer.



The diffraction gratings here are identical, and spaced so that the slits of one fall at the maxima of the previous one. As a result, atoms can propagate from A to P via either B or C. This is precisely the form of a Mach–Zehnder interferometer.

To see how this setup can detect rotation, it is easier to think of a circular interferometer, of radius R. The atoms take time $t = \pi R/v$ to propagate through each semicircular "arm", during which the entire apparatus rotates by an angle ωt. Therefore, the arms differ in effective length by $2R\omega t$, giving a phase shift

$$\Delta\phi = \frac{2\pi}{\lambda}\, 2R\omega t = \frac{4\pi}{\lambda v}\, \Omega A$$

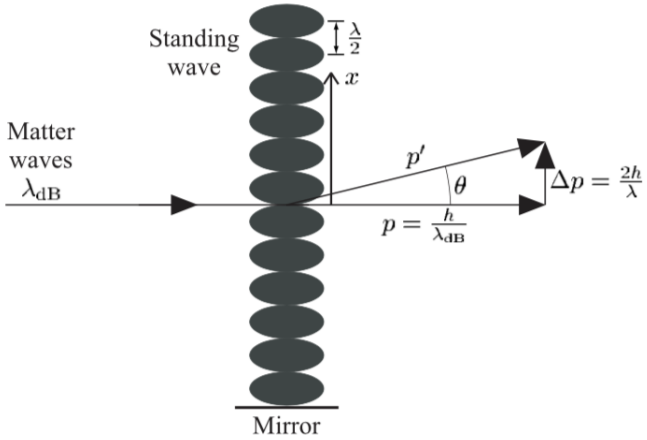
where A of the circle. A more careful derivation shows that the same result holds for any shape. (In practice, for light one would instead use a Sagnac interferometer, where light travels the same loop in opposite directions; this doubles the phase shift and cancels out some systematic effects. Of course, one can't do this for atoms since they would hit each other.)

The signal is thus parametrically much larger for atoms than for light,

$$\frac{\Delta\phi_{\text{atom}}}{\Delta\phi_{\text{light}}} \sim \frac{\lambda c}{\lambda_{\text{dB}} v} = \frac{Mc^2}{\hbar\omega}$$

from which we see the boost is due to the high mass-energy of the atoms. On the other hand, light interferometry remains competitive. Laser light is far more coherent, allowing interferometers with larger areas, which are traversed by the light many times. Also, lasers give a much higher flux of particles than atomic beams, reducing shot noise.

Example. A standing light wave can serve as a diffraction grating.



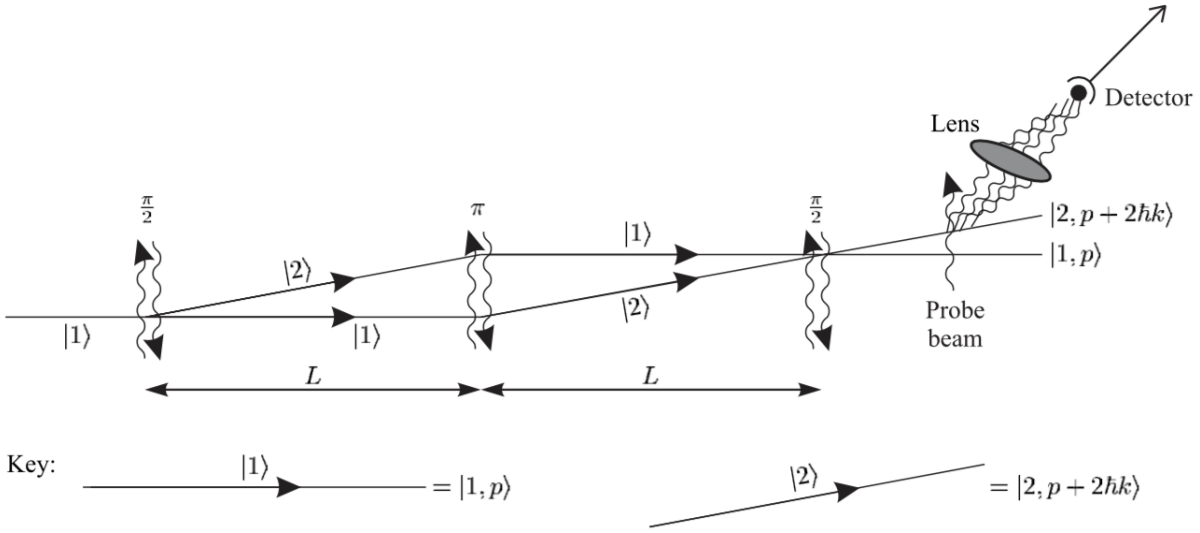
Suppose the frequency of the light is close enough to an atomic transition to cause significant interaction, but not close enough to make spontaneous emissions important. Then we can think of the light as just shifting the atomic energy levels, depending on the local field intensity. This leads to a pattern of phase shifts in the matter wave with spatial period $d = \lambda/2$, which means the grating diffracts matter waves at angles obeying

$$d \sin\theta = n\lambda_{\mathrm{dB}}$$

just like an ordinary diffraction grating. A standing wave has the advantages of transmitting all of the atoms, and of having no material that can get "clogged up" with the incident atoms. However, they have the disadvantage that the light frequency and the mirrors that form the standing wave must be very precisely stabilized.

Note that initially the atoms have no vertical momentum at all, but they end up in a superposition of having various values of vertical momentum. This momentum comes from the photons in the standing wave itself. Specifically, a standing wave is made of photons with opposite momenta. The n^{th} diffraction maximum corresponds to scattering n upward-moving photons into downward-moving photons. (Scattering into the already occupied downward-moving state is extremely likely because of its huge occupancy and the corresponding Bose enhancement; these huge occupancies also prevent these scattering events from decohering the atoms.)

Example. Interference with Raman transitions.



One can extend the idea above further. Consider two counterpropagating light waves with frequencies $\omega_1 \approx \omega_2$. If $\omega_1 - \omega_2$ is on resonance with an atomic transition (though detuned enough to prevent spontaneous emission), then by scattering a photon from one state to the other, an atom can change its internal state while simultaneously picking up an impulse of $2\hbar k$. For a $\pi/2$ pulse, the atom receives the impulse or not with equal amplitude, so it serves as a beam splitter. For a π pulse, the atom definitely receives an impulse, so it can serve as a mirror. We can also use additional π pulses to further increase the momentum difference between the two branches, thereby increasing the separation and the sensitivity.

Thus, we can build a Mach–Zehnder interferometer without needing anything "solid" at all. The two "ports" at the end correspond to the atomic states, which are read out using fluorescence from a laser tuned to a resonant transition for only one of the two states. Of course, the paths shown

above are schematic; the experiment needs to be done in freefall, so the atoms move straight up and down, accelerating downward.

Note. A wide variety of physical effects can be measured by phase shifts in an atom interferometer.

- An electric field generally causes a quadratic shift in the energy, so
$$\Delta\phi = \oint \frac{2\pi\epsilon_0\alpha}{\hbar v} \oint E^2 \, ds$$
where α is the polarizability of the state; hence atom interferometers yield sensitive measurements of polarizability.

- A magnetic flux has an effect due the usual Aharanov–Bohm effect,
$$\Delta\phi = \frac{q}{\hbar} \oint \mathbf{A} \cdot d\mathbf{s}$$
but there is also an effect of an electric field on a magnetic moment,
$$\Delta\phi = \frac{1}{\hbar c^2} \oint (\mathbf{E} \times \boldsymbol{\mu}) \cdot d\mathbf{s}$$
which is called the Aharanov–Casher phase.

- For paramagnetic atoms in a given F, M_F hyperfine sublevel, there is a phase shift due to the weak-field Zeeman effect,
$$\Delta\phi = \frac{g_F \mu_B M_F}{\hbar v} \oint B \, ds.$$

- One can also measure the phase shift due to the change in energy after a Raman transition,
$$\hbar\omega = \frac{\hbar^2 (\Delta k)^2}{2M}$$
where Δk is the impulse. Because $\hbar/M$, the Rydberg constant, and relevant mass ratios are precisely known, this gives an independent determination of α which does not require detailed QED calculations.

- The gravitational acceleration g can be measured precisely because arms of the atom interferometer with different average heights have different energy shifts due to gravitational potential energy. Specifically, if an impulse $\Delta\mathbf{k}$ is given at time $t = 0$, then the accumulated phase difference at time T is
$$\Delta\phi = \int_0^T (t\Delta\mathbf{k}) \cdot \mathbf{g} \, dt \sim (g\Delta k)T^2.$$
This has been used for precision measurements of G, and tests of the equivalence principle. One way of understanding why the sensitivity is so high is to note that the "equivalent length" of an ordinary interferometer would be $\sim cT$, which for $T \sim 1\,\mathrm{s}$ is enormous.

- One can also derive the phase shift due to rotation mentioned above by replacing $\mathbf{g}$ with the Coriolis acceleration in a rotating frame.

6.3 Ion Traps

6.4 Atomic Clocks

6.5 Rydberg Atoms