Foundations of Quantum Mechanics & Quantum Information

Dan Elton

April 23, 2012

"I think I can safely say that nobody understands quantum mechanics." – Richard P. Feynman (1965)

[When Feynman pronounced that we can never truly comprehend quantum mechanics], "he was too hasty... I think people will remove the mystery that Feynman said could never be removed... you should never say never." – Yakir Aharonov (2003)

Contents

0.1 Introduction

During the spring of 2011 I began to study quantum computation and quantum information while taking a special topics course taught by Dr. Korepin at Stony Brook University. This "book" is a series of notes I have taken to help organize what I am learning and to summarize complimentary information from different sources. These notes extend beyond quantum compuation to touch on philosophical questions I find interesting and bits of physics I have encountered elsewhere. There are various philosophical digressions on the possibility of hidden variables theories, the measurement problem and the correspondance principle. In addition to these musings, I have tried to rigorously lay out the axioms of quantum mechanics and touch on the elementary equations of quantum mechanics. Of course, these notes are not a substitute for a real textbook, but I hope that the reader will find something of interest. I have tried to focus on information which can not be found in the most popular quantum mechanics textbooks. I encourage the interested reader to check out the references listed at the end.

Chapter 1

The Foundations of Quantum Mechanics

1.1 Axioms of Quantum Mechanics

To begin I will cover the axioms of quantum mechanics. We must exercise extreme care here, because these axioms are ones on which the entire edifice of modern physics rests. (Including superstring theory!)

Postulate 1: Hilbert Space *There exists a Hilbert space $\mathcal{H}$ for every quantum system. The state of the system is given by a ray $c|\psi\rangle \in \mathcal{H}$ where $c \in \mathbb{C}$.*

Postulate 2: Operators: *For every physically measurable quantity there exists a Hermitian operator $\hat{A}$.*

Postulate 3: Measurement *Quantum measurements are described by a collection $\{M_m\}$ of measurement operators. These are operators acting on the state space of the system being measured. The index m refers to the measurement outcomes that may occur in the experiment. If the state of the quantum system is $|\psi\rangle$ immediately before the measurement, then the probability that result m occurs is given by:*

$$p(m) = \langle\psi|M_m^\dagger M_m|\psi\rangle \tag{1.1.1}$$

The state of the system after measurement is:

$$\frac{M_m|\psi\rangle}{\sqrt{\langle\psi|M_m^\dagger M_m|\psi\rangle}} \tag{1.1.2}$$

The measurement operators satisfy the **completeness relation***:*

$$\sum_i M_i^\dagger M_i = I \tag{1.1.3}$$

Postulate 4: Time development *The evolution of a closed quantum system is described by the Schrödinger equation:*

$$\hat{H}|\psi\rangle = i\hbar\frac{d}{dt}|\psi\rangle \tag{1.1.4}$$

Postulate 5: Tensor product *The state space of a composite physical system is the tensor product of the state spaces of the component physical systems.*

Postualte 6: Existence axiom *For any physical state $|\psi\rangle \in \mathcal{H}$ there exists an operator for which $|\psi\rangle$ is one of the eigenstates.*

Notes on these axioms:

Axiom 1 Equivalently the state of a system is described by a density operator ρ which is a positive operator of trace 1 defined over the Hilbert space. The density operator formalism is slightly more general, as discussed below.

Axiom 2 Shankar [10] and Nakahara [6] note at this juncture that the quantum mechanical operator for a classical observable is obtained by substituting the quatum mechanical operators $\hat{x}$ and $\hat{p}$. However there purely quantum observables (such as spin), so we keep this axiom in more general language.

Axiom 3 This formulation comes from Nielsen & Chuang[7] and can also be expressed in terms of the density operator. They use general measurement operators which may not be orthogonal like the normal projective measurement operators ($|a\rangle\langle a|$). The formula for the probability is sometimes refered to as the "Born Interpretation" (1926) or "Statistical Interpretation" for historical reasons. Since it is definitely not an interpretation it is better refered to as the "Born Rule".

Axiom 4 Equivalently, we could have used the Heisenberg equation of motion. Shankar [10] also says that H corresponds to the classical Hamiltonian with $x \to \hat{x}, p \to \hat{p}$. However, as mentioned before this is not always the case, for instance in a spin-1/2 system or in purely quantum models, so we keep the axiom general.

Axiom 5 This axiom is pretty self-explanatory but often not mentioned.

Axiom 6 This postualte appears in Nakahara [6]. It is not exactly clear to me why this postulate is needed.

There are a few other assumptions which go into the construction of quantum mechanics. One is that translations in different directions commute. This assumption is critical for deriving the **canonical commutation relations**:

$$\begin{aligned} [x_i, x_j] &= 0 \\ [p_i, p_j] &= 0 \\ [x_i, p_j] &= i\hbar\delta_{ij} \end{aligned} \tag{1.1.5}$$

Note however, that this assumption does not necessarily hold in curved spacetime.

1.1.1 Relativistic axioms

To arrive at relativistic quantum mechanics we need to add two more key axioms. We will not be discussing relativistic quantum mechanics in this book, but the axioms are given here for any aspiring unificationists. It is well accepted that the future theory of quantum gravity must subsume both the quantum mechanical axioms *and* these relativistic axioms. (A so-called "theory of everything" must also take into account numerous pieces of unexplained data such as the 19 free parameters of the Standard Model and the cosmological constant.)

1.1.2 Physical inputs

In this section I remark on some other things which are inputed into quantum mechanics.

The "Symmetrization Postulate"

1.2 Probability theory

Essential to both the calculational and philosophical aspects of quantum mechanics, but rarely discussed in much detail, are the elmentary principles of probability theory. We will review the key results of probability theory here. Probaility theory can be defined as the branch of mathematics that deals with **random variables**. To describe what a random variable is we will skip a lot of philosophical dawdle and jump to Komolgorov's axiomization of probability which has now achieved the status of orthodoxy. In his 1933 book, *Foundations of the Theory of Probability* Komologorov presented the following very general definition:

> Let Ω be a non-empty set ("the universal set"). A field (or algebra) on Ω is a set F of subsets of Ω that has Ω as a member, and that is closed under complementation with respect to Ω and union. Let P be a function from F to the real numbers obeying:
>
> 1. (Non-negativity) $P(A) \geq 0$, for all $A \in F$.
> 2. (Normalization) $P(\Omega) = 1$.
> 3. (Finite additivity) $P(A \cup B) = P(A) + P(B)$ for all $A, B \in F$ such that $A[B = \emptyset$.

Call P a probability function, and (Ω, F, P) a probability space.

1.2.1 Baye's theorem

Formally the conditional probability is defined as follows:

$$P(Y = y|X = x) \equiv \frac{P(X = x, Y = y)}{P(X = X)} \tag{1.2.1}$$

($P(X = x, Y = y)$ is the probability of x and y.) Notice that the conditional probability only deals with correlations, and does not imply causation between the processes described by X and Y. **Baye's theorem** gives a fundamental equation for a conditional probability which is designated in more compact notation as $P(B|A)$.

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \tag{1.2.2}$$

1.3 Linear Algebra

To understand Quantum Mechanics one must know some basic terminology and theorems from linear algebra. It is assumed the reader already is familiar with elementary concepts, including the definitions of a vector space, inner product, basis, orthonormality, linear independence, linear operator, matrix representation, adjoint, trace, determinant, similarity transformation, etc. A few particular points which are important for quantum computation will be reviewed here.

1.3.1 Tensor product

The **tensor product** of two vector spaces refers to the process of combining two vector spaces to form a larger one. If V is m dimensional and W is n dimensional, then $V \otimes W$ is mn dimensional. Operators can be defined on $A \otimes B$ as folows: suppose A is an operator on V and B is an operator on W. Then

$$(A \otimes B)(|v\rangle \otimes |w\rangle) \equiv A|v\rangle \otimes B|w\rangle \tag{1.3.1}$$

An explicit construction of the tensor product of operators is the so called **Kronecker product**. Basically it amounts to just a standardized way of listing (and operating on) the basis vectors of the tensor product space. It is defined as

$$(A \otimes B) = \begin{bmatrix} A_{11}B & A_{12}B & \cdots & A_{1n}B \\ A_{21}B & A_{22}B & \cdots & A_{2n}B \\ \vdots & \cdots & \ddots & \vdots \\ A_{m1}B & A_{m2}B & \cdots & A_{mn}B \end{bmatrix} \tag{1.3.2}$$

Here terms like $A_{11}B$ represent submatrices whos entries are proportional to B, with an overall proptionality constant A_{11}.

Here are two instructive (and useful) examples:

$$\begin{aligned} \sigma_z \otimes \sigma_y &= \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \otimes \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix} = \begin{bmatrix} 0 & -i & 0 & 0 \\ i & 0 & 0 & 0 \\ 0 & 0 & 0 & i \\ 0 & 0 & -i & 0 \end{bmatrix} \\ \sigma_x \otimes \sigma_z &= \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \otimes \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \\ 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{bmatrix} \end{aligned} \tag{1.3.3}$$

1.3.2 Hilbert-Schmidt inner product

A useful mathematical tool which is usually skipped over in textbooks on quantum mechanics is the **Hilbert-Schmidt inner product**. We know the space of operators is also a vector space, so it makes sense we can define an inner product on that space. There are some variations on how to do this, but the inner product that retains the properties we expect and is useful for physics is given by:

$$(A, B) \equiv \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{1}{2} A_{ij} B_{ij} = \frac{1}{2} \operatorname{tr}(A^\dagger B) \tag{1.3.4}$$

(Technically the Hilbert-Schmidt inner product shouldn't have a factor of 1/2, but in quantum mechanics it is useful because it preserves normalization, for instance, if you are decomposing an arbitrary 2x2 matrix into Pauli matrices.) As a side note, the **Frobenius norm**, also called the Hilbert-Schmidt norm or Euclidian matrix norm is similarily defined as:

$$||A||_F = \sqrt{\mathrm{tr}(A^\dagger A)} \tag{1.3.5}$$

1.3.3 Important theorems in linear algebra

Any Unitary matrix U can be written as

$$U = e^{iH} \tag{1.3.6}$$

where H is a Hermitian matrix.

1.4 The Density Matrix

The density matrix / density operator formalism was developed by John Von Neumann in 1927 and independently by Landau around the same time. Quoting Wikipedia, "The motivation that inspired Landau was the impossibility of describing a subsystem of a composite quantum system by a state vector. On the other hand, von Neumann introduced the density matrix in order to develop both quantum statistical mechanics and a theory of quantum measurements." A thorough discussion of the density matrix can be found in the book by Sakurai[9]. The density matrix is central to most research in quantum mechanics outside of particle theory, so it is somewhat suprising that it is not covered much in undergraduate quantum mechanics courses. My feeling is that this is so because of an underlying tendancy to focus on reductionist thought in physics. This tendancy is not hard to understand, nor is it unreasonable – reductionism has been remarkably sucessful in the past 300 years. However, things are starting to change now, as we have a good grasp of fundamental building blocks, a lot of the exciting research today is in understanding what happens when they come together. The density matrix, by it's nature, becomes of use only when considering a system of interest as a subsystem of a larger system, or when considering many quantum systems coupled together. Both of these cases are the opposite of reduction. Because the density matrix is becoming the tool of choice in both small "open quantum systems" coupled to the environment, in statistical and chemical physics and even in calculating abstract properties of quantum fields, my guess is that it will be emphasized more in the future. Before introducing the density matrix we should distinguish between what are sometimes called **quantum probability** and **classical probability**. Classical probability deals with situations where we have a lack of knowledge. For instance, we have a bag of 100 red marbles and 100 blue marbles. If a marble is selected randomly, the probability it is red is 1/2. Quantum probability is more intrinsic – even in theory we cannot know what color we will get. With quanutm probability, a single marble can be in a "coherent linear superposition" of states – $\frac{1}{\sqrt{2}}(|R\rangle + |B\rangle$. Furthermore, only with quanutm probability is there quantum interference, leading to the possibility that $P(A+B) \neq P(A) + P(B)$.

The density matrix is a way of handling a large ensemble of quantum systems $|psi_i\rangle$, where the classical probility of selecting a particular type of sub-system is given by w_i. The definition of the density matrix is:

$$\rho \equiv \sum_i w_i |\psi_i\rangle\langle\psi_i| \tag{1.4.1}$$

Because $\sum_i w_i = 1$, $\mathrm{tr}(\rho) = 1$. It can be also shown that ρ must be a positive operator, meaning it's eigenvalues are all positive. In fact, it can be proved that any matrix that satisfies these two conditions, $\mathrm{tr}(\rho) = 1$ and positivity, is a density matrix describing some quantum system. We next define the **ensemble average**, which gives the average value of a variable over an ensemble of quantum systems:

$$[A] \equiv \sum_i w_i \langle\psi_i|A|\psi_i\rangle \tag{1.4.2}$$

Now we note a compact way of computing this average:

$$\begin{aligned} \langle A\rangle &= \sum_i w_i \langle\psi_i|M|\psi_i\rangle \\ &= \sum_i \sum_i w_i \langle\psi_i||j\rangle\langle j|M|\psi_i\rangle \\ &= \sum_i \sum_j w_i \langle j|M|\psi_i\rangle\langle\psi_i||j\rangle \\ &= \mathrm{tr}(M\rho) \end{aligned} \tag{1.4.3}$$

A pure state is just one ensemble, ie. $\rho = |\psi\rangle\langle\psi|$. It follows that for a pure state, $\rho^2 = \rho$ and therefore $Tr(\rho^2) = 1$ in addition to $Tr(\rho) = 1$. For mixed states, $(\rho^2)1$.

1.4.1 Time Evolution of the Density Matrix

The time evolution of the density operator follows from Schrödinger's equation.

$$\begin{aligned} i\hbar\frac{\partial\rho}{\partial t} &= (i\hbar\frac{\partial\rho}{\partial t}|\psi\rangle)\langle\psi| + |\psi\rangle(i\hbar\frac{\partial\rho}{\partial t}\langle\psi| \\ &= H|\psi\rangle\langle\psi| - |\psi\rangle\langle\psi|H \\ &= -[\rho, H] \end{aligned} \tag{1.4.4}$$

A pure state remains pure under time evolution, which can be verified by checking that the $\mathrm{tr}(\rho^2)$ property is conserved.

$$\begin{aligned} \frac{d}{dt}\mathrm{tr}(\rho^2) &= \mathrm{tr}(2\rho\frac{d\rho}{dt}) \\ &= \frac{2i}{\hbar}\mathrm{tr}(2\rho(\rho H - H\rho)) \\ &= \frac{2i}{\hbar}[\mathrm{tr}(\rho\rho H) - \mathrm{tr}(H\rho\rho)] = 0 \end{aligned} \tag{1.4.5}$$

1.4.2 Ambiguity of the density matrix

It turns out that two states which are related by a unitary transformation will have the same density matrix. A rigorous proof of this fact is given in Nielsen & Chuang, pg. 104.[7]

1.4.3 Reduced Density Matrix

Consider a Hilbert space $\mathcal{H} = \mathcal{H}_A \oplus \mathcal{H}_B$. The **reduced density matrix** for system A is defined as:

$$\rho^A \equiv \mathrm{tr}_B(\rho^{AB}) \tag{1.4.6}$$

tr_B is known as the **partial trace** of ρ^{AB} over B. The partial trace is defined as :

$$\mathrm{tr}_B(|a_1\rangle\langle a_1||b_1\rangle\langle b_2|) \equiv |a_1\rangle\langle a_1|\,\mathrm{tr}(|b_1\rangle\langle b_2|) \tag{1.4.7}$$

As we expect, if our state is just a tensor product a density matrix ρ_A in $\mathcal{H}_A$ and a density matrix ρ_B in $\mathcal{H}_B$ then

$$\rho^A = \mathrm{tr}_B(\rho_A\rho_B) = \rho_A\,\mathrm{tr}(\rho_B) = \rho_A \tag{1.4.8}$$

A less trivial example is the bell state.

1.4.4 Entropy & Entanglement of a Density Matrix

The entanglement entropy, or "Von Neumann Entropy" is defined as:

$$\sigma \equiv -\,\mathrm{tr}(\rho\ln\rho) \tag{1.4.9}$$

For a pure state, the only eigenvalue is 1 and the entropy is 0. For a maximally disordered state, $\sigma = \sum_i \frac{1}{n}\ln\frac{1}{n} = \ln n$. Note that with this definition, the additive property of entropy is preserved by the natural log. It turns out that this definition of entropy is also a good measure of the entanglement of a quantum system.

1.4.5 Continiuum Form of the Density Matrix

So far we have only been discussing quantum systems of finite dimensionality. For continuous systems (ie. wavefunctions describging particles moving in potentials), the denisity matrix becomes a product of wavefunctions. This form of the density matrix recieves little or no attention in major textbooks on quanutm mechanics even though technically it is no less fundamental than the discrete form. Landau & Lifshitz [5] defines it as follows:

$$\rho(x,x') = \int \Psi(q,x)\Psi^*(q,x')dq \tag{1.4.10}$$

Here, q represents extra coordinates of the system which are "not being considered" in the measurements we will perform. What is not explicit in this definition is what is analogous to the w_i in the discrete case. We must assume that some or all of the q's play the role of the $w_i's$. Indeed, Landau notes that for a pure state, $\rho(x,x') = \Psi(x)\Psi^*(x')$, no integration over q required. The (equivalent) definition in Sakurai[9] is:

$$\begin{aligned}\rho &= \langle x''|\left(\sum_i w_i\left|\alpha^{(i)}\right\rangle\left\langle\alpha^{(i)}\right|\right)|x'\rangle \\ &= \sum_i w_i\psi_i(x'')\psi_i^*(x')\end{aligned} \tag{1.4.11}$$

1.4.6 Example: Infinite Square Well

As an example let us consider a particle sitting in a infinite square well with the state $\psi(x) = \sqrt{\frac{2}{\pi}}\sin(x)$, $x \in [0, \pi]$. The density matrix is:

$$\rho = \frac{2}{\pi}\sin x' \sin x$$

Now let us break the square well in half into two subspaces, $\mathcal{H}_A \otimes \mathcal{H}_B$. Somewhat suprisingly, now we can calculate entropy and entanglement of these subspaces, even though the entire system is pure (and has zero entropy and entanglement). The key difference now is that in each subspace we no longer know if there is a particle or not. Thus, we must express each sub-Hilbert space in the basis $(|0\rangle + |1\rangle) \otimes \psi(x)$ where $|0\rangle$ represents no particle and $|1\rangle$ represents one particle. Equivalently, you can think of this as a way of breaking up the wave function into two functions.

$$\psi = |0, 0\rangle \otimes |1, \psi(x)\rangle + |1, \psi(x)\rangle \otimes |0, 0\rangle \tag{1.4.13}$$

Or, compressing notation and ignoring normalization for now,

$$\psi = |01\rangle + |10\rangle \tag{1.4.14}$$

Now we can find the reduced density matrix for $\mathcal{H}_A$. First we rewrite ρ:

$$\begin{aligned} \rho &= (|01\rangle + |10\rangle)(\langle 01| + \langle 10|) \\ \rho &= |01\rangle\langle 01| + |10\rangle\langle 10| + |10\rangle\langle 01| + |01\rangle + \langle 10| \end{aligned} \tag{1.4.15}$$

The trace of $|10\rangle\langle 01|$ is zero since $\mathrm{tr}(|10\rangle\langle 01|) = \mathrm{tr}(\langle 01|10\rangle) = 0$. We are left with

$$\begin{aligned} \rho_A &= \frac{2}{\pi}\sin x \sin y |1\rangle\langle 1| + \int_{\pi/2}^{\pi} \frac{2}{\pi}\sin(x)^2 dx |0\rangle\langle 0| \\ \rho_A &= \frac{2}{\pi}\sin x \sin y |1\rangle\langle 1| + \frac{1}{2}|0\rangle\langle 0| \end{aligned} \tag{1.4.16}$$

To calculate the entropy of this subsystem we must know the eigenvalues of ρ_A. By inspection, it is clear that $|0\rangle$ is an eigenvector with eigenvalue $\frac{1}{2}$. We know the sum of the eigenvalues must be 1, so the other eigenvalue is $\frac{1}{2}$. (Alternativly we could try $f(x)|1\rangle$ and work out the corresponding inner product, which is an integral.)

$$S = -\mathrm{tr}(\rho \ln(\rho)) = -\frac{1}{2}\ln\frac{1}{2} - \frac{1}{2}\ln\frac{1}{2} = \ln(2) \tag{1.4.17}$$

This result could have been anticipated because there are two possibilities for measurement – we find the particle on the left or the right. Still, it is suprising that the entropy of a given subsystem can be non-zero while the entire system has zero entropy.

1.4.7 Gleason's theorem

Throughout this chapter we have remarked how the axioms of quantum mechanics can be recast in terms of the density matrix.

The version given in Wikipedia is as follows: *For a Hilbert space of dimension 3 or greater, the only possible measure of the probability of the state associated with a particular linear subspace a of the Hilbert space will have the form* $\mathrm{tr}(P(a)\rho)$*, the trace of the operator product of the projection operator* $P(a)$ *and the density matrix* ρ *for the system.*

Gleason's theorem has a deep significance.

1.5 Schmidt decomposition

Suppose $|\Phi\rangle$ is a pure state of a composite system AB. Then there exist orthonormal states $|i_A\rangle$ for system A, and orthonormal states $|i_B\rangle$ for system B such that

$|\Phi\rangle = \sum_i \lambda_i |i_A\rangle|i_B\rangle$where λ_i are non-negative real numbers satisfying $\sum_i \lambda_i^2 = 1$. This new expression is

Closely related to the Scmidt decomposition is the procedure of **purification**.

1.6 Purification

1.7 The Bloch Sphere

The block sphere is a representation of a spin 1/2 system.

It is based on the observation that, for spin 1/2, an arbitrary density matrix can be expressed in terms of the Pauli matrices and the identity, which form a basis for the space of 2x2 matrices.

$$\rho = \frac{\hat{I} + \vec{r} \cdot \vec{\sigma}}{2}. \tag{1.7.1}$$

The factor of 1/2 ensures that $\mathrm{tr}(\rho) = 1$ (the Pauli matrices are traceless and don't contribute). The term "sphere" is a misnomer, technically it is a ball since $0 \leq r \leq 1$. The surface of the sphere corresponds to pure states, which can be proven by showing $\rho^2 = \rho$:

$$\begin{aligned} \rho^2 &= \frac{1}{4}[\hat{I} + 2\vec{r} \cdot \vec{\sigma} + (\vec{r} \cdot \vec{\sigma})^2] \\ &= \frac{1}{4}[\hat{I} + 2\vec{r} \cdot \vec{\sigma} + \hat{I}] \\ &= \frac{1}{2}[\hat{I} + \vec{r} \cdot \vec{\sigma}] = \rho \end{aligned} \tag{1.7.2}$$

Likewise, the center of the sphere corresponds to the maximally disordered state.

1.8 The Three Pictures

There are three standard formalisms for writing operators and state kets: The Schrödinger picture, the Heisenberg picture and the interaction (Dirac) picture. We will not describe these pictures in detail but assume the reader is already familiar with them. They are related as follows:

1.9 Quantum Dynamics

To characterize a quantum system, the fundamental problem of quantum mechanics is to find the Hamiltonian that describes the system. Once the Hamiltonian is known the dynamics are given by Schrödinger's equation:

$$H|\Psi\rangle = i\hbar \frac{\partial}{\partial t}|\Psi\rangle \tag{1.9.1}$$

The Baker-Hausdorff lemma is

Solving the resulting differential equation is in most cases non-trivial, and there are many approaches. The formal solution is given by

$$|\Psi(t)\rangle = e^{\frac{i}{\hbar}Ht}|\Psi(0)\rangle \tag{1.9.2}$$

In the Schrödinger picture and

Let us now consider a general time-dependent potential. Let H_0 be the time independent part of the Hamiltonian. Then analysis becomes easier if we choose to work the **interaction picture**. The interaction picture state ket is related to the Schrödinger state ket as follows:

$$|\psi, t_0; t\rangle_I = e^{iH_0 t/\hbar}|\psi, t_0; t\rangle_S \tag{1.9.3}$$

Here the notation $|\psi, t_0, t\rangle$ is understood to mean that the system is in the state $|\psi\rangle$ at t_0 and then evolves in time. Observables transform as follows:

$$A_I \equiv e^{iH_0 t/\hbar} A_S e^{-iH_0 t/\hbar} \tag{1.9.4}$$

The Interaction Picture, or Dirac Picture, is an intermediary between the Schrödinger Picture and the Heisenberg Picture.(see pg 336 of Sakurai [9]). The time-evolution operator in the interaction picture is defined as

$$|\psi, t_0; t\rangle_I = U_I(t, t_0)|\psi, t_0; t_0\rangle_I \tag{1.9.5}$$

$$U_I(t, t_0) i\hbar \frac{d}{dt} U_I(t, t_0) = V_I(t) U_I(t, t_0) \tag{1.9.6}$$

We must solve this differential equation with the intial condition $U_I(t_0, t_0) = 1$. The solution is called the **Dyson Series** after Freeman J. Dyson, who applied this method in QED. We can rewrite 1.9.6 as an integral equation:

$$U_I(t, t_0) = 1 - \frac{i}{\hbar}\int_{t_0}^{t} V_I(t') U_I(t', t_0) dt' \tag{1.9.7}$$

We now solve this by iteration.

(1.9.8)

*** PUT EQUATIONS HERE ****

1.10 The Path Integral Formulation

Let us revisit our equation for time development:

$$|\Psi(t')\rangle = e^{\frac{i}{\hbar}H(t'-t)}|\Psi(t)\rangle \tag{1.10.1}$$

The path integral formulation provides an entirely new way of looking at quantum mechanics.

Transfering to the Heisenberg picture, we say that the path integral gives us thte overlap between eigenstates of a position operator. In otherwords, let $Q(t)$ be a time-dependent

position operator and let $|q,t\rangle$ be a Heisenberg picture state such that $Q(t)|q,t\rangle = q|q,t\rangle$. The path integral allows us to compute the quantity:

$$U(q'',t'';q',t') = \langle q'',t''|q',t'\rangle \tag{1.10.2}$$

This quantity is extremely useful in quantum field theory, where it is the basis of the S-matrix (scattering matrix).

There are several ways to arrive at the path integral. I will take what is probably the most mathematically transparent route but certaintly not the most intuitive. More conceptually illustrative depictions of the path integral are given in the class book by Feynman & Hibbs.[1]

We start by rewriting the expotential $e^{i-H(t''-t))/\hbar}$ as a product:

U(q",t";q',t') $=\lim_{n\to\infty} \langle q''(t'')|(1-\frac{iH\delta t}{\hbar})^n|q'(t')\rangle$
(1.10.3)

We now insert unity $I = \int dq(t)|q(t)\rangle\langle q(t)|$ between each of the terms in the product, yielding

U(q",t";q',t') $=\lim_{n\to\infty} \langle q''(t'')|\prod_{i=1}^{n}\int dq_i(t)|q_i\rangle\langle q_i|aq''(t'')(1-\frac{iH\delta t}{\hbar})^n|q'(t')\rangle|q-1\rangle\langle q-1|$
(1.10.4)

Now comes probably the most non-obvious step in this line of derivation. We want to show that the following equality holds:

$$\lim_{n\to\infty}\prod_{i=1}^{n-1}\left[1-ih(p_i,q_i)\frac{t''-t'}{n\hbar}\right. = \exp\left[\lim_{n\to\infty}\frac{-i(t''-t')}{n\hbar}\sum_{i=1}^{n-1}h(p_i,q_i)\right] \tag{1.10.5}$$

Take the natural logarithm of both sides.

$$\ln\lim_{n\to\infty}\prod_{i=1}^{n-1}\left[1-ih(p_i,q_i)\frac{t''-t'}{n\hbar}\right] = \ln\exp\left[\lim_{n\to\infty}\frac{-i(t''-t')}{n\hbar}\sum_{i=1}^{n-1}h(p_i,q_i)\right] \tag{1.10.6}$$

The left hand side is

$$\lim_{n\to\infty}\sum_{i=1}^{n-1}\ln\left[1-ih(p_i,q_i)\frac{t''-t'}{n\hbar}\right] \tag{1.10.7}$$

Now expand the natural logirthm in the $n\to\infty$ limit using the Taylor series:

$$\ln(x) \approx -x - \frac{x^2}{2} - \frac{x^3}{3}\cdots \tag{1.10.8}$$

$$= \lim_{n\to\infty}\sum_{i=1}^{n-1}\left[-ih(p_i,q_i)\frac{t''-t'}{n\hbar} + \frac{1}{2}h(p_i,q_i)^2(\frac{t''-t'}{n\hbar})^2 - \cdots\right] \tag{1.10.9}$$

The right hand side is

$$= \sum_{i=1}^{n-1} \left[\lim_{n\to\infty} \frac{-i(t''-t')}{n\hbar} h(p_i, q_i) \right] \tag{1.10.10}$$

So we see that the two expressions are equivalent to order $1/n$ and thus equivalent in the limit $n \to \infty$.

1.10.1 Path integral for the free particle

The equation for the path integral becomes:

$$U(q'', t''; q', t') = \lim_{n\to\infty} \left(\frac{m}{2\pi i\hbar\delta t}\right)^{n/2} \int_{-\infty}^{\infty} \prod_{i=1}^{n-1} dq_i \exp\left(\frac{im}{2\hbar\delta t} \sum_{j=1}^{n-1} (q_{j+1} - q_j)^2\right) \tag{1.10.11}$$

Define $y_i = \sqrt{\frac{im}{2\hbar\delta t}} q_i$, with $dq_i = \sqrt{\frac{2\hbar\delta t}{im}} dy_i$. Then

$$U(q'', t''; q', t') = \lim_{n\to\infty} \left(\frac{m}{2\pi i\hbar\delta t}\right)^{n/2} \left(\frac{2\hbar\delta t}{im}\right)^{\frac{n-1}{2}} \int_{-\infty}^{\infty} \prod_{i=1}^{n-1} dy_i \exp\left(\sum_{j=1}^{n-1} (y_{j+1} - y_j)^2\right)$$

(1.10.12)

This expression looks intimidating, but it can actually be done systematically. It helps to look at the y_2 integral first, which is

$$\begin{aligned} &= \int_{-\infty}^{\infty} \exp\left((y_2 - y_1)^2 + (y_3 - y_2)^2\right) \\ &= \int_{-\infty}^{\infty} \exp\left(2y_2^2 - 2(y_1 + y_3)y_2 + y_1^3 + y_3^2\right) \\ &= \sqrt{\frac{\pi}{2}} e^{y_1^2 + y_3^2} e^{-\frac{4(y_1+y_3)^2}{8}} \\ &= \sqrt{\frac{\pi}{2}} e^{\frac{y_1^2}{2} + \frac{y_3^2}{2} - \frac{y_1+y_3}{2}} \\ &= \sqrt{\frac{\pi}{2}} e^{\frac{1}{2}(y_3 - y_1)^2} \end{aligned} \tag{1.10.13}$$

The y_3 integral becomes

$$i\sqrt{\frac{\pi}{2}} \int_{-\infty}^{\infty} \exp\left((y_4 - y_3)^2 + \frac{1}{2}(y_3 - y_1)^2\right) \tag{1.10.14}$$

This is similiar to the y_2 integral, except for the pesky factor of $\frac{1}{2}$. Plodding through this integration yields

$$i\sqrt{\frac{\pi}{2}}i\sqrt{\frac{2\pi}{3}}e^{\frac{1}{3}(y_4-y_1)^2} \quad (1.10.15)$$

We can see a patern here. The next integral will yield a prefactor of $i\sqrt{\frac{3\pi}{4}}$ and a $\frac{1}{4}$ in the exponent. After $n-1$ integrations, we will arrive at

$$(i\sqrt{\pi})^{n-1}\sqrt{\frac{1}{n-1}}e^{\frac{1}{n-1}(y_{n-1}-y_1)^2} \quad (1.10.16)$$

The net result, after doing all n integrals, is:[1]

$$\sqrt{\frac{m}{i\hbar\delta t}}\left(\frac{1}{\pi}\right)^{n/2}(-i)^{\frac{2n-1}{2}}(i\pi)^{\frac{n}{2}}\sqrt{\frac{1}{n}}\exp\left(\frac{1}{n}(y_n-y_1)^2\right) \quad (1.10.17)$$

After replacing y_i with q_i and noting that $\delta tn = t'' - t'$, we arrive at the expression for the free particle propagator:

$$U(t'',q';q',t') = \sqrt{\frac{m}{2\pi i\hbar(t''-t')}}\exp\left(\frac{-m}{2i\hbar}\frac{(q''-q')^2}{(t''-t')}\right) \quad (1.10.18)$$

We can also evaluate the Lagrangian $L = m\dot{q}^2/2 + k\dot{q}$ in a very similiar manner. The analysis proceeds as before, but now we have a term $k(q_{j+1} - q_j)$ in the exponential as well. As you can see, the sum $\sum_j^n (q_{j+1} - q_j)$ will just be $q_n - q_1$. The result becomes.

Note that the phase part of the propagator contains the classical action for the free particle. In Feynman and Hibbs it is shown that this is always the case as long as the Lagrangian contains terms not higher than quadratic. In such cases, the exact phase factor can be written down immediatly. The coefficient out front is harder to compute, and can only be found exactly in special cases. The path interal is of limited utility in non-relativistic quantum mechanics, although it can provide some insight and makes the classical limit easy to see. In recent decades the path integral has become of great utility in quantum field theory. In fact, for some non-abelian gauge theories the only way that things can be calculated is through the path integral approach.

1.11 The Uncertainty Principle

1.11.1 The Heisenberg Uncertainty Relation

The derivation of the uncertainty principle is a tad bit technical and can be found in almost all books on quantum mechanics. (for instance, Shankar [10] Chapter 9, Sakurai [9] Section 1.4, etc). The **dispersion** or **variance** or **mean square deviation** of an observable A is defined as:

$$\sigma_A^2 \equiv \left\langle (A - \langle A\rangle)^2 \right\rangle = \left\langle A^2 \right\rangle - \langle A\rangle^2 \quad (1.11.1)$$

[1] There is undoubtably a mistake here with the factors of i, however I have not bothered to correct it as such net phase factors are usually inconsequential. The correct result will have a factor of $\sqrt{1/i} = \sqrt{-i}$ out front, but as pointed out in the revised addition of Feynman & Hibbs [1], this standard form is ambigious because we don't know which branch of the square root to take. A full analysis of the phase factor involves specifying the branch and is derived in a research paper cited in the revised version of Feynman & Hibbs.

The general result for two observables $\hat{A}$ and $\hat{B}$ is:

$$\sigma_A^2 \sigma_B^2 \geq \frac{1}{4} |\langle [A, B] \rangle|^2 \tag{1.11.2}$$

Leading to

$$\sigma_x^2 \sigma_p^2 \geq \frac{\hbar^2}{4} \tag{1.11.3}$$

$$\sigma_x \sigma_p \geq \frac{\hbar}{2} \tag{1.11.4}$$

1.11.2 The energy-time uncertainty principle

The energy time uncertainty principle is distinct from the Heisenberg Uncertainty relation and no less fundamental. It is given by:

$$\Delta E \Delta t \geq \hbar/2 \tag{1.11.5}$$

The reason for this uncertainty principle comes from the wave nature of the wavefunction. Heuristically it is described in Shankar [10], pg. 245. Eingenstates of energy have a time dependance factor $e^{-iEt/\hbar}$, ie. a definte energy is associated with a definite frequency $\omega = E/\hbar$. However, only a wave train which is infinitely long in time has a definite frequency. (There is some frequency spread in the Fourier transform of the wavefunction, and hence a spread in energy) Thus, a system which has been around only a finite time cannot be associated with a pure frequency or a definite energy. The significant aspect of this is that *energy need not always be conserved*, in fact, in any finite-time physical process, there is always some degree of energy non-conservation. To me, this is perhaps the most startling feature of quantum mechanics, in a way even more startling then the statistical nature of the theory. It means, in effect, that *in principle* anything is possible! To derive this important relation, we need to use the time-dependent perturbation theory mentioned above.

The implications of the energy-time uncertainty principle are far reaching. For instance, in section 5.9 of Sakurai[9] it is derived (using 2nd order time-dependent perturbation theory) how the "decay width" Γ and the "decay lifetime" τ of an unstable state are related. The relation is a direct manifestation of the time-energy uncertainty principle:

$$\tau \Gamma = \hbar \tag{1.11.6}$$

This means the uncertainty in energy is inversely proportional to the lifetime of the state. This principle is applicable to atomic physics, but perhaps more dramatically, to particle physics. It means that shortly lived particles have a distribution of energies and via $E = mc^2$ a corresponding distribution in masses. The fact that an elementary particle can have many different masses is strange indeed!

In light of all this, one might wonder how many ultra-high energy cosmic rays are due to statistical fluctuations in energy.

1.12 Symmetries

1.12.1 Lie groups and algebras

1.12.2 Time reversal symmetry

1.13 Second Quantization

1.13.1 The Fock Space

The procedure of "Second Quantization" was developed to handle quantum systems with many particles. The state of the system is described by a vector in "Fock Space" which looks like this:

$$|n_1, n_2, \ldots, n_i, \ldots\rangle \tag{1.13.1}$$

Here n_i specifies the number of particles with eigenvalue k_i for some operator. The idea that we can describe the state of a quantum system with such a state vector seems natural enough, but implies some delicate assumptions.

We now introduce two special economical notations. First is the notation for the ground state (no particles)

$$|0, 0, \ldots, 0, \ldots\rangle \equiv |\mathbf{0}\rangle \tag{1.13.2}$$

Next is the notatio for a state with just one particle with eigenvalue k_i

$$|0, 0, \ldots, 0, 1, 0, \ldots\rangle \equiv |k_i\rangle \tag{1.13.3}$$

We define a creation operator ("field operator") $a_i^\dagger$ that increases the number of particles in the state with eigenvalue k_i by one

$$a_i^\dagger |n_1, n_2, \ldots, n_i, \ldots\rangle = C|n_1, n_2, \ldots, n_i + 1, \ldots\rangle \tag{1.13.4}$$

The corresponding anihilation operator a_i is postulated to satisfy the following conditions:

$$\begin{aligned} a_i|n_1, n_2, \ldots, n_i, \ldots\rangle &= C^*|n_1, n_2, \ldots, n_i - 1, \ldots\rangle \\ a_i|\mathbf{0}\rangle &= 0 \\ a_i|k_j\rangle &= \delta_{ij}|\mathbf{0}\rangle \end{aligned} \tag{1.13.5}$$

The number operator is defined as

$$N = \sum_i a_i^\dagger a_i \tag{1.13.6}$$

For bosons, the comm

Chapter 2

Measurement

2.1 The Measurement Problem

The measurement problem actually predates Schrödinger's cat and starts with Einstein's lesser known powder keg. On August 8, 1935 Einstein sent a letter to Schrödinger in which he described an unstable powder keg. Refering to the quantum mechanical description of the keg, he remarked:

> After a year the psi-function then describes a sort of blend of not-yet and alreadyexploded systems. Through no art of interpretation can this psi-function be turned into an adequate description of a real state of affairs; in reality there is just no intermediary between exploded and not-exploded.

On September 19, 1935 Schrödinger replied to Einstein

> I have constructed an example very similar to your exploding powder keg..

and went on to describe the famous cat which bears his name. Later that month Einstein replied:

> Your cat shows that we are in complete agreement concerning our assessment of the character of the current theory.... A psi-function that contains the living as well as the dead cat cannot be taken as a description of the real state of affairs.

Indeed, both Einstein *and* Schrödinger were metaphysical realists. Both the bomb and the cat where meant to dramatically illustate an disparity in quantum mechanics between unitary evolution and state reduction.
From axiom 3 we know that a measurement causes the state to collapse to an eigenstate. As a historical note, we mention that this collapse was not present in the orginal quantum theory. The collapse axiom was added after the famous Compton-Simons experiment in 1926. Collapse is now very well established experimentally, and can be tested by making two consecutive measurements which should give the same result. In fact, if a system is "constantly measured", it cannot evolve – this is the basis of the quantum Zeno effect.
As mentioned, in quantum mechanics there are two distinct processes: unitary time evolution described by axiom 4 and state reduction (collapse) described by axiom 3:

Unitary Evolution	**State reduction (collapse)**
Deterministic	Non-deterministic
Continuous	Discontinuous
Time-reversible	Not time-reversible*
Thermodynamically reversible	Not thermodynamically reversible*

*in general

But why should the measuring device, which is just a composition of many microscopic quantum systems, have such different behavior when compared to the microscopic quantum systems of which it is composed? Can state reduction be subsumed into unitary evolution? If not, then what exactly constitutes a measurement – ie. what are the necessary and sufficient conditions for collapse to occur? This question is the "measurement problem".

We should note that there are interpretations of quantum mechanics which do not have collapse (ie. the many world interpretation). This possibility will be discussed in more detail later.

2.2 Measurement theory

2.2.1 Von Neumann measurement formalism

Von Neumann, in his famous work *The Mathematical Foundations of Quantum Mechanics* (1932), struggled to formalize wavefunction collapse mathematically and ultimently was forced to conclude that consciousness causes collapse. However, he came up with a useful formalism which describes the quantum mechanical interaction between the measurement device and the system being measured. We first break up the world into three components: I the microscopic system to be measured, II the (macroscopic) measuring device, and III the observer. The Von Neuman formalism (and all all measurement formalisms) describe the interactions between I and II only. For conveince, and without loss of generality, the meausring device is usually thought of a "a pointer" with some (quantum mechanical) position and momentum.

To describe measurment Von Neumann first introduces the **ideal measurement hamiltonian**:

$$H_M = q(t)PA \tag{2.2.1}$$

Here, $q(t)$ is a coupling function which describes the interaction (which is compactly supported and normalized to 1 for convience). P is the conjugate momentum operator of pointer variable Q and A is the operator for whatever is being measured. During the measurement the Hamiltonian of the complete system is given by:

$$H = H_I + H_{II} + q(t)PA \tag{2.2.2}$$

We assume that in the (brief) time window of measurement, $[t_1, t_2]$ the dynamics from H_I and H_{II} can be ignored. Then the time evolution is given by:

$$|\psi(t_2)\rangle = \exp(-\frac{i}{\hbar}\int_{t_1}^{t_2} q(t)PAdt)|\psi(t_1)\rangle \tag{2.2.3}$$

Without loss of generality we assume the system is in a pure state with eigenvalue a.

$$|\psi(t_2)\rangle = \exp(-\frac{i}{\hbar}Pa)|\psi(t_1)\rangle \tag{2.2.4}$$

This is simply a translation operator which translates the pointer a distance a, proportional to the quantity measured. Notice that system I is not disturbed by the measurement - ie. it remains in it's pure state without disruption. Which this formalism is very useful, it is important to remember that ideal measurements are just that - ideal!

2.2.2 Projective (Strong) measurements

In postulate three we already noted that measurements are described by measurement operators which are projectors on the Hilbert space. Projective measurements are a special case of more general measurement where the projectors are all orthogonal.

Projective measurements can be used to calculate average values:

$$\begin{aligned} < M > &= \sum_m mp(m) \\ &= \langle\psi| \sum_i mP(m)|\psi\rangle \\ &= \langle\psi|M|\psi\rangle \end{aligned} \tag{2.2.5}$$

2.2.3 Quantum Non-Demolition (QND) Mesurements

A ideal measurement that can be realized in the laboratory is called a quantum non-demolition measurement. The telltale property of a QND measurement is that repeated measurments yield the same result. It is a common misconception to think that QND measurements do not effect the system being measured. They do collapse the wavefunction, but do not have any "back-action" on the system from the measuring device. For example a system in a pure state will remain in a pure state with no perturbation to the system.

Examples of non-demolition measurements include: [**?**]

- An Stern-Gerlach apparatus which does block the beam. (a somewhat idealized notion, in my opinion)
- measurement of the the electromagnetic field energy stored inside a cavity by dtermining the radiation pressure on a moving piston. (Braginsky & Khalili, 1992)
- detection of the presence of a photon in a cavity by its effect on the phase of an atom's superposition state. (Nogues, et. al. 1999)
- Measurement of a qubit state by it's effect on the frequency of a driven microwave resonator.

QND's are very difficult to achieve in the laboratory. Most measurements are demlotion measurements. Obvious examples of demolition measurements are:

- Position measurements using light.
- Detectors which destroy the incoming particle.
- any measurement where the observable does not commute with the Hamiltonian of the system.

2.2.4 Interaction free measurements

The concept of an interaction free measurement (IFM) was first theorized by Mauritius Renninger in 1953 and popularized by Elitzur and Vaidman in 1993 with thier proposal for a "bomb testing device". The "Elitzur Vlaidman bomb tester" is the canonical example, and many explanations of it can be found online. I will give a brief review of it here.

Suppose we have some very sensitive bombs. The trigger is so sensitive, that if we place a mirror on the trigger, the trigger will be activated if a single photon bounces off of it. Some of the bombs are duds. Are goal is to determine which bombs are duds, without actually destroying any of the actual bombs. The miraculous solution was named by *New Scientist* magazine as one of the seven wonders of the quantum world.

The solution uses a Mach-Zender style interferometer. If there is no bomb, the photon's wave function travels in the lower and upper path and self-interferes.

(section under construction)

What does this interaction-free measurement tell us?

2.2.5 POVM measurements

A strange property of quantum mechanics is that non-orthogonal states cannot be distinguished. To illustrate, suppose that Alice selects (prepares) a state $|\psi_i\rangle$ from a set of possible states and gives it to Bob. Bob then must make a measurement to determine what state it is. If Bob has a set of projective measurements he can perform which span the space $M_i = \langle\psi_i||\psi_i\rangle$ with the same basis, then he can apply these measurements and he will find the state $|\psi_i\rangle$ with certainty. But now suppose that Alice's states are not all orthogonal. Then any some states may share a component. When Bob measures and finds a result he can't be sure what state Alice choose. This is a fundamental problem in quantum computing. To help distinguish states, we can try to use more general non-orthogonal measurements called POVM measurements which, as we will see, can allow better (but still not perfect) distinguishing of states.

"POVM" is a term from functional analysis which stands for "Positive Operator Valued Measure". A collection of measurement operators $M_i^\dagger M_i$ is refered to as "a POVM". [1] As before, they obey the completeness relation:

$$\sum_i M_i^\dagger M_i = I \tag{2.2.6}$$

however here the operators *need not be orthogonal*.

Because the operators in a POVM need not be orthogonal, POVM are not always repeatable – they do not always collapse the wavefunction to a pure state which can be immediatly measured again, yielding the same result. If ρ' is subjected to the same measurement, the new state is

$$\rho'' = \frac{M_i \rho' M_i^\dagger}{\mathrm{tr}(M_i \rho' M_i^\dagger)} = \frac{M_i M_i \rho M_i^\dagger M_i^\dagger}{\mathrm{tr}(M_i M_i \rho M_i^\dagger M_i^\dagger)} \tag{2.2.7}$$

which is equal to ρ' iff $M_i^2 = M_i$ that is, if the POVM reduces to a projective measurement.

[1] As the story goes, these operators used to be known as "superoperators" until the advent of supersymmetry when a string theorist ran into a quantum information theorist. To avoid confustion, the new name was adopted.

2.2.6 POVM Example

This example is taken from Nielsen Chuang, pg. 92.[7] Suppose Alice gives Bob a qubit which is prepared in one of two states: either $|\psi_1\rangle = |0\rangle$ or $|\psi_2\rangle = (|0\rangle + |1\rangle)/\sqrt{2}$. Consider a POVM which consists of three measurements:

$$\begin{aligned} E_1 &= \frac{\sqrt{2}}{1+\sqrt{2}}|1\rangle\langle 1| \\ E_2 &= \frac{\sqrt{2}}{1+\sqrt{2}}\frac{(|0\rangle - \langle 1|)(|0\rangle - \langle 1|)}{2} \\ E_3 &= I - E_1 - E_2 \end{aligned} \tag{2.2.8}$$

Suppose that Bob applies these measurements and gets E_1. Then he knows for sure that the states must have been $|\psi_2\rangle$. Like wise, if he receives the result E_1, then he knows the state must have been $|\psi_1\rangle = |0\rangle$. Sometimes, however he will recieve the outcome E_3, and in that case he will not be sure which state he had. The advantage of this scheme is that he can distinguish the two states with certainty at least some of the time.

2.3 The Strong-Weak measurement continuum

(under construction) Appart from the special QND measurements, most measurements must lie on a spectrum between weak and strong (ideal).

Weak measurement was theorized by Aharonov, Albert, Vaidman in 1987. Theory revised by M. Duck, P. M. Stevenson, and E. C. G. Sudarshan (1989) (and other articles) Hulet, Ritchie, Story, (1991) First experimental realization. Hosten Kwiat (2008) - used weak measurements to measure the Spin Hall Effect for photons. Splitting of light beam 1 Angstrom. Dixon et.al (2009) measured angular deflection of a light beam with the precision of a hairs breadth at the distance of the moon.

2.4 Two State Vector Formalism

Chapter 3

Hidden Variables

3.1 The Bell Inequality

John Bell has been called "the deepest thinker in the foundations of quantum mechanics". Up until 1964 many physicists still believed that quantum mechanics was incomplete and that hidden variables were needed. It was that year that John Bell proved that any *local* hidden variables theory is incompatable with quantum mechanics. To prove this, Bell considered the orginal EPR setup of two decay products moving in opposite directions, but with a slight modification - the observers become free to choose what axis they will measure spin along. (There are of course many subsequent variations in the formulation of the Bell's inequality, this particlar description follows the exceptionally lucid discussion in Griffiths.[3]) The first observer measures along a unit vector $\hat{\mathbf{a}}$, the other along a unit vector $\hat{\mathbf{b}}$. For simplicity, let's asume the possible values of spin are $+1, -1$. Bell proposed that the observers calculate the average of the product of the two spin values, which we denote $A(\hat{\mathbf{a}}, \hat{\mathbf{b}})$. If the detectors are aligned in the same direction, the product is always -1 and so is the average. So,

$$A(\hat{\mathbf{a}}, \hat{\mathbf{a}}) = -1 \tag{3.1.1}$$

Quantum mechanics predicts that

$$A(\hat{\mathbf{a}}, \hat{\mathbf{b}}) = -\hat{\mathbf{a}} \cdot \hat{\mathbf{b}} \tag{3.1.2}$$

Let us quickly verify this equation. The wavefunction of the two particle system is given by:

$$|\psi\rangle = \frac{1}{\sqrt{2}}(|\uparrow\downarrow\rangle - |\downarrow\uparrow\rangle) \tag{3.1.3}$$

The minus sign is critical here, since this is a singlet state. We are interested in:

$$\left\langle (\mathbf{S_1} \cdot \hat{\mathbf{a}})(\mathbf{S_2} \cdot \hat{\mathbf{b}}) \right\rangle = \frac{1}{2}(\langle\uparrow\downarrow| - \langle\downarrow\uparrow|(\mathbf{S_1} \cdot \hat{\mathbf{a}})(\mathbf{S_2} \cdot \hat{\mathbf{b}})(|\uparrow\downarrow\rangle - |\downarrow\uparrow\rangle \tag{3.1.4}$$

Let us align our coordinate system so that $\hat{\mathbf{a}}$ is in the z direction and $\hat{\mathbf{b}}$ lies in the x-z plane. Then we have:

Bell showed that for *any* local hidden variables theory, the following inequality holds:

$$|A($$

The quantum mechanical prediction can be shown to be incompatable with this inequality, for instance by looking $\hat{\mathbf{a}}$ and $\hat{\mathbf{b}}$ at a 45 degree angle. Then $A(\hat{\mathbf{a}}, \mathbf{c}) = A(\hat{\mathbf{b}}, \mathbf{c}) = -.707$ and there is a dramatic inequality:

$$.707 \not\leq 1 - .707 = .293 \tag{3.1.7}$$

3.2 Aftermath

Bell's inequality showed that nature, if it is to be considered real, must be nonlocal. This sounds shocking, but the degree of shock depends on how one defines "locality". Certaintly, this is not a violation of the law that no matter or energy travel faster than light implicit in Einstein's theory, which is what the term "locality" refers to in discssions of Quatum Field Theory. Yet, it is a violation of Einstein's definition of locality – that an event at one point should not super-luminally influence another event. DJ Griffiths does a good job of explaining why this is not as shocking as it may seem. It requires looking closely at what we describe as an "influence". There are many "influences" in nature which travel faster than light. For instance, a spacecraft may create a shadow which move at hyper-luminal speeds, given that the screen it is projected on is far enough away. This shadow carries no energy, yet, if a chain of detectors are placed along the path of the shadow one would notice very distinct correlations (or "influences") propagating through the chain at faster than light speed. The key point is that these correlations are not *causual*. The Bell experiment is the same way – when Alice measures his spin, she can not cause it to be a certain value – rather, quantum mechanics causes the measurement outcome. Likewise Bob does the same, and it is not until afterwards when they compare their data that remarkable correlations are noticed between the data. In summary, the movement of material particles, (atoms, photons, etc) is still limited to the speed of light. The "speed of correlations" though, can be faster than light.

3.2.1 Loopholes

The reason there have been so many different Bell test experiments is that many of the early experiments were plauged with subtle "loopholes".

3.2.2 Superdeterminism

Although "Bell's inequality" is usually thought to be the end of local hidden variables, Bell himself pointed out that there is still the lingering possibility of"superdeterminism".[11] Bell described this possibility in a BBC interview in 1985:

> *There is a way to escape the inference of superluminal speeds and spooky action at a distance. But it involves absolute determinism in the universe, the complete absence of free will. Suppose the world is super-deterministic, with not just inanimate nature running on behind-the-scenes clockwork, but with our behavior, including our belief that we are free to choose to do one experiment rather than another, absolutely predetermined, including the "decision" by the experimenter to carry out one set of measurements rather than another, the difficulty disappears. There is no need for a faster than light signal to tell particle A what measurement has been carried out on particle B, because the universe, including particle A, already "knows" what that measurement, and its outcome, will be.*

Some thought makes this objection seem very unlikely. The reason it is unlikely is not that it violates "free will", most scientists who are literate on the subject of free will realize it to be a rather useless metaphysical concept and are more than willing to give it up.. but that is another story. Superdeterminism seems unlikely when we suppose the decision of what spin direction to measure is made by some complex system which is dependent on a large number of tiny effects (for instance, a human brain or an electronic random number generator). It seems extremely unlikely that hidden variables would be sensitive to all these tiny effects at Alice's device and then conspire miraculously with other tiny effects at Bob's device to produce the measurement correlations we observe. Because of this very little has been written about superdeterminism. Another reason is that it is like an existance proof - we see that local hidden variables are still possible, but we have absolutely no indication of how to construct such a theory. Perhaps the only serious physicist who has stated superdeterminism cannot be ignored is Gerald 't Hooft.

Chapter 4

The Correspondance Principle

4.1 The general philosophy of the Correspondance Principle

The correspondance principle is a principle at the foundations of science which predaes quantum mechanics. Classical-quantum correspondance is today an entire field of study. Our understanding of how the classical world emerges from the quantum has many small holes, which researchers are filling in, and many deep mysteries which may require new ways of thinking about nature to solve. Overall, progress is slow in this field, for two reasons. The first may be tht many of the best young physicists are drawn into the more "fundamental research". The thrust of physics, and science in general (as well as much of Western though) has been reductionist – in other words, we are most interested in the fundamental constituents of matter and how they act – the rest is considered "filling in details".

The second reason why progress is slow in quantum - classical correspondance is that it is, in general, extremely difficult to formulate it in precise mathematical terms. Of course, fundamental physics (particullarily string theory) suffers from the same issue. However, the field of exactly solvable models in quantum statistical mechanics gives us a rare glimse at how classical behavior emerges from micrscopic quantum interactions. The most celebrated example is discovery of a phase transition in the exact solution to the 2D Ising model. One can see precisly, in precise elegant analytical terms, how the collective and purely macroscopic phenomina of phase transition comes out of the interplay of interacting spins. In this model, we are given a glimse at emergence laid bare. It is perhaps the only example where we understand *perfectly* how a macroscopic phenomina emerge from microscopic phenomina.

To get full understanding of quantum mechanics, one must get an idea of how it reduces to classical mechanics. The transition to classical behangor ocurs when the number of particles in the system becomse very large or when the mass of particles become large.

One idea we should dispose of is the idea that classical mechanics is simply the limit of quantum mechanics when "$\hbar$ goes to zero". While in mathematical sense this procedure "usually" works, it is not a good way to approach the problem of classical-quantum corresponance. Classical-quantum correspondance is difficult precisely because $\hbar$ is "not" zero in the real world. In the words of Dr. Korepin, "There is no knob in the lab where we can adjust $\hbar$". Furthermore, the trick of taking the limit of $\hbar$ going to zero may result

in a severe loss of information. For instance, the band theory of solids, energy density and volume of ice, superconductivity, and many other phenomina "are" quantum phenomina that are manifest on a macroscopic scale. They would be missed if one simply took $\hbar$ to zero.

4.2 Ehrenfest's theorem

Ehrenfest's theorem is the analog of Newton's 2nd law in quantum mechanics. It can be derived straightforwardly from the Heisenberg equation of motion:

$$\frac{dA^H}{dt} = \frac{1}{i\hbar}\left[A^H, H\right] \tag{4.2.1}$$

and the Hamiltonian:

$$H = \frac{p^2}{2m} + V(\vec{x}) \tag{4.2.2}$$

Then,

$$\begin{aligned} \frac{dp_i}{dt} &= -\frac{\partial}{\partial x_i}V(\vec{x}) \\ \frac{dx_i}{dt} &= -\frac{p}{m} \\ \frac{d^2x_i}{dt^2} &= \frac{1}{i\hbar}\left[\frac{dx_i}{dt}, H\right] = \frac{1}{m}\frac{dp_i}{dt} = -\frac{1}{m}\frac{\partial}{\partial x_i}V(\vec{x}) \end{aligned} \tag{4.2.3}$$

Yielding, in vectorial form,

$$m\frac{d^2\vec{x}}{dt^2} = -\nabla V(\vec{x}) \tag{4.2.4}$$

Or, taking expectation values of both sides, we get an equation which makes sense in either picture:

$$m\frac{d^2}{dt^2}\langle\vec{x}\rangle = -\langle\nabla V(\vec{x})\rangle \tag{4.2.5}$$

ie, the center of a quantum mechanical wave moves classically.

4.3 Coherent States of the Harmonic Oscillator

When most students learn about the harmonic oscillator as an undergraduate, they learn about the wavefunctions of the stationary states. In the large n limit, the stationary states have the probability distribution we expect classically – low probability in the center, where the oscillator spends less time, and higher probability near the classical turning points, where the oscillator slows down. Yet, these states are not the ones usually found in nature – they are non-dynamical, after all. If we take a classical oscillator, such as a pendulum and make it very, very small, we do not get a stationary state. In general, we get a coherent state. The coherent states are states that are dynamic in time – the wavefunction moves with a sinusoidal motion which nicely reduces to the classical case when the mass becomes large. Remarkably, these states also minimize uncertainty.

Coherent states were first discovered by Schrödinger in 1926 while he was searching for states which would satisfy the correspondence principle. He suspected that such states would minimize the uncertainty relation. I do not know what motivated him to search for such uncertainty-minimizing states; perhaps his physical intuition was telling him that

classical states should be well localized in position and momentum space.
Let us first remind ourselves that the ground state of the harmonic oscillator is a minimum uncertainty state - thus it is a coherent state. It is also a state which has a classical correspondence – as an oscillator at rest. To show this, we utilize the following well known relations:

$$\begin{aligned} x &= \sqrt{\frac{\hbar}{2m\omega}}(a + a^\dagger) \\ p &= i\sqrt{\frac{m\omega\hbar}{2}}(a - a^\dagger) \end{aligned} \tag{4.3.1}$$

Since clearly $(\langle x\rangle)^2 = 0$ and $(\langle p\rangle)^2 = 0$, we thus have

$$\sigma_x^2\sigma_p^2 = \left\langle x^2\right\rangle \left\langle p^2\right\rangle = \langle 0|\frac{\hbar}{2m\omega}(a + a^\dagger)^2\frac{(-1)m\omega\hbar}{2}(a - a^\dagger)^2|0\rangle \tag{4.3.2}$$

$$\begin{aligned} \langle 0|(a + a^\dagger)^2|0\rangle &= \langle 0|aa^\dagger|0\rangle = 1 \\ \langle 0|(a - a^\dagger)^2|0\rangle &= \langle 0| - aa^\dagger|0\rangle = -1 \end{aligned} \tag{4.3.3}$$

So, clearly

$$\sigma_x^2\sigma_p^2 = \frac{\hbar}{2m\omega}\frac{m\omega\hbar}{2} = \frac{\hbar^2}{4} \tag{4.3.4}$$

If we repeat this analysis with an arbitrary state $|n\rangle$ we will find that the $a^\dagger a$ term will begin to contribute and that they will no longer be uncertaintly minimizing states. The key feature that gave us uncertainty minimization was that $a|\psi\rangle = 0$. This clues us into the fact that we can find other uncertainty minimizing states by looking for states which are eigenstates of the anihilation operator $\hat{a}|\alpha\rangle = \alpha|\alpha\rangle$. This property is what defines coherent states.

In the energy ($|n\rangle$) basis, the coherent states can be derived to have the following form:

$$|\alpha\rangle = \sum_n^\infty |n\rangle\langle n||\alpha\rangle$$

Since

$$\begin{aligned} |n\rangle &= \frac{a^{\dagger n}}{n!}|0\rangle \\ \langle n| &= \langle 0|\frac{a^n}{n!} \end{aligned} \tag{4.3.6}$$

And hence,

$$|\alpha\rangle = \sum_n^\infty \frac{\alpha^n}{n!}|n\rangle\langle 0|n\rangle \tag{4.3.7}$$

The term $\langle 0|n\rangle$ is a constant which will go away after normalization. A quick normalization calculation yields:

$$|\alpha\rangle = e^{-\frac{1}{2}|\alpha|^2}\sum_n^\infty \frac{\alpha^n}{n!}|n\rangle \tag{4.3.8}$$

Before moving on, I will mention a few technical points about coherent states. The first is that there is a completeness relation for coherent states:

$$\int d^2\alpha |\alpha\rangle\langle\alpha| = \pi \tag{4.3.9}$$

which shows that they span the Hilbert space. However, the coherent states are not orthogonal:

$$|\langle\alpha|\beta\rangle|^2 = e^{-|\alpha-\beta|^2} \tag{4.3.10}$$

The coherent states are said to form an "overcomplete basis". This fact can be traced back to the fact that $\hat{a}$ is not Hermitian.

Now, let us consider the time evolution of coherent states:

$$\begin{aligned} |\alpha,t\rangle &= e^{-iHt/\hbar}|\alpha,0\rangle \\ &= e^{-\frac{1}{2}\alpha(0)^2}\sum_n^\infty \frac{\alpha^n}{\sqrt{n}} e^{-i\hbar\omega(n+1/2)/\hbar}|n\rangle \\ &= e^{-\frac{1}{2}\alpha^2(0)}\sum_n^\infty \frac{\alpha^n}{\sqrt{n}} e^{-i\hbar\omega(n+1/2)/\hbar}\frac{(a^\dagger)^n}{\sqrt{n}}|0\rangle \\ &= e^{-\frac{1}{2}\alpha(0)^2}e^{-i\omega/2}e^{\alpha a^\dagger e^{-i\omega t}}|0\rangle \end{aligned} \tag{4.3.11}$$

We see that, apart from the phase factor of $e^{-i\omega t/2}$, we have another coherent state, with time dependent eigenvalue $\alpha(t) = e^{-i\omega t}\alpha(0)$. Thus, the coherent state remains coherent under time evolution.

Now note that

$$|\alpha,t\rangle = e^{-i\omega t/2}\big|e^{i\omega t}\alpha(0)\big\rangle \longrightarrow \frac{d}{dt}\alpha = -i\omega\alpha(t) \tag{4.3.12}$$

We are now in a position to take the classical limit. The expectation values are defined as :

$$\begin{aligned} x(t) &= \langle\alpha(t)|x|\alpha(t)\rangle \\ p(t) &= \langle\alpha(t)|p|\alpha(t)\rangle \end{aligned} \tag{4.3.13}$$

$$\begin{aligned} x(t) &= \sqrt{\frac{\hbar}{2m\omega}}(\alpha(t)+\alpha^*(t)) = \sqrt{\frac{\hbar}{2m\omega}}\,\mathrm{Re}(\alpha(t)) \\ p(t) &= i\sqrt{\frac{m\omega\hbar}{2}}(\alpha(t)-\alpha^*(t)) = i\sqrt{\frac{m\omega\hbar}{2}}2\,\mathrm{Im}(\alpha(t)) \end{aligned} \tag{4.3.14}$$

But note that equation 4.3.12 implies that:

$$\begin{aligned} \frac{d}{dt}\,\mathrm{Re}(\alpha(t)) &= \omega\,\mathrm{Im}(\alpha(t)) \\ \frac{d}{dt}\,\mathrm{Im}(\alpha(t)) &= -\omega\,\mathrm{Re}(\alpha(t)) \end{aligned} \tag{4.3.15}$$

After a bit of algebra (left to the reader) we arive at the classical equations of motion:

$$\begin{aligned} p(t) &= m\frac{d}{dt}x(t) \\ \frac{d}{dt}p(t) &= -m\omega^2 x(t) \end{aligned} \tag{4.3.16}$$

It is quite amazing to see how the classical equations of motion emerge from the quantum formalism.

In quantum field theory, (particularily, quantum electrodynamics) coherent states of a field exist and correspond classically to electromagnetic waves.

4.4 The WKBJ / Quasiclassical Approximation

There are several ways to arrive at the WKB Approximation. Normally the WKB equation is derived using an ansatz imvolving an amplitude and a phase: $\psi(x) = A(x)e^{i\phi(x)}$. Here we take a more direct approach, based on an expansion of the wavefunction in $\hbar$, which is given as problem 8.2 in Griffiths.[3], pg. 320. Motivated by the free particle wavefunction $\psi(x) = A\exp(\pm ipx/\hbar)$ we express the wavefunction as:

$$\psi(x) = e^{\frac{i}{\hbar}f(x)} \tag{4.4.1}$$

Where $f(x)$ is some complex function. There is no loss of generality here – any non-zero function can be expressed in this form. Schrödinger's equation can be re-arranged as follows:

$$\begin{aligned} -\frac{\hbar^2}{2m}\psi'' + V(x)\psi &= E\psi \\ \psi'' = -\frac{p^2}{\hbar^2} \quad p(x) &\equiv \sqrt{2m[E - V(x)]} \end{aligned} \tag{4.4.2}$$

We plug 4.4.1 into this equation to get, after some differentiation,:

$$i\hbar f'' - (f')^2 + p^2 = 0 \tag{4.4.3}$$

Now we expand $f(x)$ as a power series in $\hbar$:

$$f(x) = f_0(x) + \hbar f_1(x) + \hbar^2 f_2(x) + \cdots \tag{4.4.4}$$

$$\begin{aligned} i\hbar f_0''(x) + i\hbar^2 f_1''(x) + i\hbar^3 f_2''(x) + \cdots - p^2 = (f_0'(x) + \hbar f_1'(x) + 2f_2'(x)^{-} + \cdots) \\ \times (f_0'(x) + \hbar f_1'(x) + 2f_2'(x)^{-} + \cdots) \end{aligned} \tag{4.4.5}$$

$$p^2 + \hbar i f_0'' + \mathcal{O}(\hbar^2) = f_0'^2 + 2\hbar f_1' f_0' + \mathcal{O}(\hbar^2) \tag{4.4.6}$$

And as usual, we compare terms with equal power of $\hbar$:

$$\begin{aligned} f_0'^2 &= p^2 \\ i f_0'' &= 2 f_0' f_1' \\ f_0' = \pm p(x) &\Longrightarrow \boxed{f_0 = \pm \int_0^x p(x)dx} \\ -f_0'' &= 2p(x) f_1' \\ f_1' = \frac{ip'}{2p} &\Longrightarrow \boxed{f_1 = \int \frac{ip(x)'}{2p(x)}dx} \\ f_1 &= i\frac{1}{2}\ln(p) \\ f = = &\pm \int_0^x p(x)dx + i\hbar \ln(\sqrt{p}) + \cdots \end{aligned} \tag{4.4.7}$$

Therefore, by expanding $f(x)$ to first order in $\hbar$ we obtain the WKB expansion formula for the wavefunction:

$$\psi(x) \cong \frac{C_1}{\sqrt{p(x)}} e^{\frac{i}{\hbar}\int_0^x p(x)dx} + \frac{C_2}{\sqrt{p(x)}} e^{-\frac{i}{\hbar}\int_0^x p(x)dx} \tag{4.4.8}$$

Out of curiosity, let us calculate the term of order $\hbar^2$. We arrive at the equation:

$$\begin{aligned} i\hbar^2 f_1'' &= 2f_0' f_2' \hbar^2 + f_1'^2 \hbar^2 \\ f_2' &= \frac{i f_1'}{2f_0'} - \frac{f_1'^2}{2f_0'} \\ &= \pm(-\frac{p''}{4p^2} + \frac{p'^2}{p^3} + \frac{p'^2}{8p^3}) \\ f_2 &= \pm(-\frac{p'}{4p^2} + \int_0^x \frac{p'^2}{8p^3} dx) \end{aligned} \tag{4.4.9}$$

$$\psi(x) \cong \frac{C_1}{\sqrt{p(x)}} e^{\frac{i}{\hbar}\left(\int_0^x p(x)dx - \frac{p'}{4p^2} + \int_0^x p'^2 8p^3 dx\right)} + \text{c.c.} \tag{4.4.10}$$

Or, we can expand part of the exponent as follows:

$$\begin{aligned} \psi(x) &\cong e^{\frac{i}{\hbar}(f_0+f_1}(1 - i\hbar f_2) \\ \psi(x) &\cong \frac{C_1}{\sqrt{p(x)}} e^{\frac{i}{\hbar}\int_0^x p(x)ds} \left[1 + \frac{p'}{4p^2} - \int_0^x p'^2 8p^3 dx\right] + \text{c.c.} \end{aligned} \tag{4.4.11}$$

Note that the force equals:

$$F = -\frac{dU}{dx} = \frac{p'p}{m} \tag{4.4.12}$$

$$\psi(x) \cong \frac{C_1}{\sqrt{p(x)}} e^{\frac{i}{\hbar}\int_0^x p(x)ds} \left[1 - \frac{mF}{4p^4} - \int_0^x \frac{mF^2}{8p^5} dx\right] + \text{c.c.}$$

This is not a very nice expression to look at it but it is helpful in determining when the WKB approximation is valid. For instance, we see that the "force" (change in potential) must be adequately small and must go to zero at $\pm\infty$. Conceptually, we see that our ansatz is somewhat like a wavefunction of a free particle with varying wavelength, so we expect the approximation to be valid when the wavelength $(1/k)$ is much smaller than the scale over which the potential varies. In other words, for distances spanning several wavelenghts we can regard the potential as constant. More formally, let us look at 4.4.3. We see that the first order term does not overwhelm the zero-th order term provided that

$$\hbar\,|f''| \ll |f'|^2 \tag{4.4.14}$$

This inequality is always true as $\hbar \to 0$. Since $f'(x) \cong p(x) = \hbar k(x)$, this is equivalent to saying that

$$|k'(x)| \ll |k^2(x)| \tag{4.4.15}$$

Or, since $k = \frac{2\pi}{\lambda} = \frac{1}{\hbar}\sqrt{2m(E - V(x))}$ we arrive at

$$\lambda \ll \frac{4\pi|E - V(x)|}{|dV/dx|} \tag{4.4.16}$$

Which confirms that WKB is applicable in the limit of small wavelength. We can also rewrite 4.4.15 as :

$$\begin{aligned} \left|\frac{k'}{k^2}\right| &\ll 1 \\ \left|\left(\frac{1}{k}\right)'\right| &\ll 1 \\ |\lambda'| &\ll 1 \end{aligned} \tag{4.4.17}$$

which shows that in the WKB approximation, the wavelength must change slowly.

4.5 Connection to the Hamilton-Jacobi equation of Classical Mechanics

What does $f(x)$ correspond to classically? The *time dependent* wavefunction has the form:

$$\Psi \propto e^{\frac{i}{\hbar}(f(x)-Et)} \tag{4.5.1}$$

This is the same as **Hamilton's characteristic function** $W(x$. In classical mechanics, the action is also called "Hamilton's principle function". When a classical system has a time-independent Hamiltonian, the action can be seperated as follows:

$$S(x,t) = W(x) - Et \tag{4.5.2}$$

4.6 The Density Matrix and Quantum Statistical Mechanics

The density matrix is useful because real world experiments are usually done using ensembles of quantum systems (say, a beam of electrons, a beam of photons or a small collection of atoms). When the number of particles in the ensemble becomes very large, we enter the realm of quantum statistical mechanics and the results of classical statistical mechanics. Thus there is a direct connection between the density matrix and statistical mechanics which we will now elucidate. We already took a major step in this direction by defining the (Von Neumann) entropy in terms of the density matrix. It turns out that the Von Neumann entropy is equivalent to the Boltzmann entropy per particle one learns in statistical mechanics, given that we multiply by k.

$$S_{Boltzmann} = k\S \tag{4.6.1}$$

4.6.1 Simple model of dephasing

Roughly speaking, there are two effects on a quantum system due to coupling to the environment - energy loss and decoherence.

As was discussed in Chapter 1, the time evolution of the density matrix is given by :

$$i\hbar\dot{\rho} = [H, \rho] \tag{4.6.2}$$

We break our Hamiltonian into system, environment, and interaction parts:

$$H = H_s + H_e + H_{int}$$

The simplist possible system we can study is a two level system, which can be described by the Hamiltonian

$$H_x = a\sigma_z \tag{4.6.4}$$

The interaction term will be easiest to deal with if it is "factorable", ie. if :

$$H_{int} = -f\{\lambda\}\sigma_z \tag{4.6.5}$$

An example of a system described by this model is a spin-1/2 particle in a fluctuating magnetic field. (in such case $a = \mu_B < B_z >$ and $f = -\mu_B B_z(t)$) Plugging into the Heisenberg equation of motion for σ_z shows us that $\dot{\sigma_z} = 0$.

Now assume that the system is described by an arbitrary density matrix:

(4.6.6)

4.6.2 Fluctuation-dissipation theorem

The celebrated fluctuation-dissipation theorem:

Chapter 5

Quantum paradoxes and experiments

> "...quantum mechanical phenomina do not occur in a Hilbert space, they occur in a laboratory." -Asher Peres

5.1 The double slit with a (delayed) quantum eraser

In this section I will discuss an important experiment called the **Quantum Eraser experiment**. The description is mainly based off the website `http://grad.physics.sunysb.edu/~amarch/`. The experiment was first performed in 1982 paper by Scully and Drhl and has been repeated several times since. The classic double experiment has been scrutinized many times by many great physicists. Feynman once said that all the magic and mystery of quantum mechanics can be found in this experiment.

The quantum eraser experiment shows emphatically that the Heisenberg uncertainty principle cannot due solely by the measuring device disturbing the system. This point is often confused, because (non-ideal) measuring devices do, of course, disturb the system, and because historically Heisenberg motivated the uncertainty principle by thinking about position measurements using photons (The so called "Heisenberg Microscope"). Any explanation of the uncertainty principle using disturbance (as often appears in popular physics articles) is wrong – the uncertainty principle is a innate property of any physical system, which comes out of the axioms of quantum mechanics and has no known deeper explanation.

The quantum eraser experiment uses entangled photons. They are entangled in orthogonal directions (a maximally entangled Bell state). The entangled photons are created via a process called spontaneous parametric down conversion. Light from an argon laser (351.1 nm) enters a nonlinear crystal called beta-barium borate. Each photon that enters the crystal is converted into two photons of longer wavelength (702.2 nm) which are entangled. Conveniently, the photons can be made to go off in different directions.

One photon goes to a detector, the other photon goes through a double slit. We wil call the detector photon "p" and the double slit photon "s". Another detector at the double slit measures the displacement of the photon after passing through. If the apparatus is run like this, a normal interference pattern is observed. The possibility for a quantum eraser comes by introducing a means of "tagging" the lower photon to determine which slit it

went through. A quarter wave plate is placed in front of each slit. One wave plate turns the photon into right circularly polarized, the other turns it into left circularly polarized. Now by measuring the polarization of "p" we can find which slit "s" went through, since they are entangled. (If the polarization of "p" is in the x (y) direction, the polarization of "s" before going through the slit must have been in the y (x) direction, thus we know it will be right (left) circularly polarized if it goes through slit 1 and left (right) circularly polarized if it goes through slit 2). The interference pattern disappears. Note that it is not necessary to disturb the "s" photon as it goes through the slits. In fact, it is also not necessary to actually measure the polarization of "p". Merely being able to is enough to destroy the interference pattern. This proves without a doubt that the uncertainty principle cannot be explained merely by "the measuring device disturbing the system". One may wonder if the quarter wave plates alone introduce a disturbance to the interference pattern. First, we note that the polarization of the photons has no effect on the interference pattern in the normal double slit – one can send in light of any polarization and the pattern persists. It is true that if we place the quarter wave plates in front of the slits, with unentangled, polarized light, the interference pattern will be destroyed. But as the next section shows, we can restore the interference pattern even with the quarter wave plates in place by the use of a "quantum eraser".

The quantum eraser is introduced by placing a polarizer in the "p" beam. This polarizer polarizes photons into a mixture of x and y directions. Now we can no longer know with certainty what the initial polarization of "p" was. The interference pattern returns. The polarizer "erases" the which-way information. The obvious question is, "how does the "s" photon know the polarizer is there?". It might be many miles away, yet the result will be the same. This is the mystery of entanglement.

Perhaps the "p" photon sends a message to the "s" photon when it encounters the polarizer. Can we test if this is the case? Indeed we can, by performing what is called "delayed quantum erasure". We extend the distance that "p" travels so that by the time it reaches the polarizer, the "s" photon has already passed through the double slit and landed on the detector. Miraculously, the interference patter appears! Now even an instantaneous message from "p" to "s" when it reaches the polarizer is pointless. Somehow, the system knows that the which way information will be lost, even before photon "p" reaches the polarizer. This flies against our instincts about local interactions, and proves emphatically that entanglement is a non-local phenomena.

But we can go one step further, with the "delayed choice quantum eraser" experiment performed by Yoon-Ho Kim, R. Yu, S.P. Kulik, Y.H. Shih, and Marlan O. Scully in 2000. Their experimental setup is a bit more complicated, and involves beam splitters to redirect the photons. Their setup is equivalent to introducing an beam splitter in the path of "p" which either redirects it towards a polarizer (eraser) or towards a detector *after* the "s" photon has been detected. When the output of the detector is observed the interference pattern is washed out. But by looking at the subset of photons who's entangled partner did not go through the eraser, the interference pattern is found. If one looks at the subset of photons whose partner went through the eraser, there is no interference pattern. Somehow, the "s" photon knows what direction the 'p" photon will travel, normally considered a random quantum mechanical process. Some have interpreted this as "retrocausality", however there is not really any physical justification for such a notion.

5.2 Schrödinger's Cat

5.3 The quantum zeno effect

Chapter 6

Information Theory

"All things physical are information-theoretic in origin and this is a participatory universe ... Observer participancy gives rise to information; and information gives rise to physics." J. A. Wheeler (1990), (1994)

6.1 Classical Information Theory

This is a very brief summary of the key definitions and results of classical information theory.

6.1.1 Shannon entropy

Let X denote a probability distribution over a set $\{x\}$. The **Shannon entropy** associated with X is defined as:

$$H(X) \equiv -\sum_x p_x \log p_x \tag{6.1.1}$$

Technical note: in all discussions which follow, we define $0 \log 0 \equiv 0$. This is mathematically justified as $\lim_{x \to \infty} x \log x = 0$. The "log" in this equation, and equations which follow, is usually taken in base 2. Changing to base e or base 10 only introduces a scaling factor (through the rule $\log_a(x) = \frac{\log_b(x)}{\log_b(a)}$). The fundamental unit of information is the "bit" which is defined by considering a variable which takes two values with equal likelihood, leading to entropy $-\log_2(1/2) = \log_2(2) = 1$. If natural logarithms are taken, the unit is called a "nat", and if base 10 is used, the unit is called a "dit" or "Hartely".

6.1.2 Self entropy

The **self entropy** of a message m is defined as

$$I(m) = -\log(p(m)) \tag{6.1.2}$$

where $p(m)$ is the probability of receiving message m out of the space of all possible messages. With this definition, we see that the Shannon entropy is the average self-entropy over the space of possible messages.

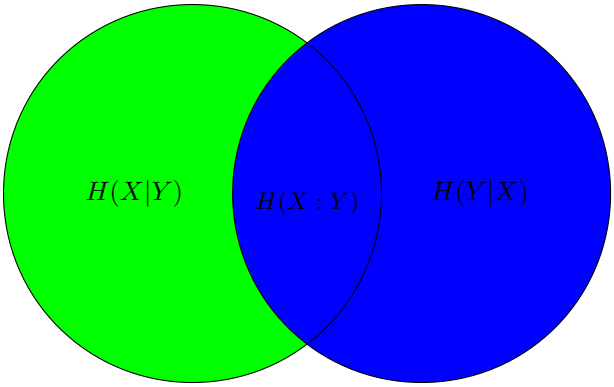


Figure 6.1: A so-called "Information diagram", which shows the relationships between different measures of information. Mutual information $H(X:Y)$, Conditional entropies $H(X|Y)$ and $H(Y|X)$. The joint entropy $H(X,Y)$ encompasses the entire diagram.

6.1.3 Other entropy measures

Suppose $p(x)$ and $q(x)$ are two probability distributions on the same set, x. The **relative entropy** is defined as:

$$H(p(x)||q(x)) \equiv \sum_x \log \frac{p(x)}{q(x)} \equiv -H(p(x)) - \sum_x p(x) \log q(x) \quad (6.1.3)$$

Naturally, for two distributions X and Y over the sets x and y the **joint entropy** is defined as:

$$H(X,Y) \equiv -\sum_{x,y} p(x,y) \log p(x,y) \quad (6.1.4)$$

The **conditional entropy** of X conditional on knowing the value of Y is:

$$H(X|Y) \equiv H(X,Y) - H(Y) \quad (6.1.5)$$

Finally, we define the **mutual entropy** or "mutual information", which measures how much information X and Y have in common. Suppose we add the information of X, $H(X)$ to the information of Y. Information which is common to both will be counted twice. It makes sense that the definition subtracts off the joint information $H(X,Y)$ to correct for this double counting:

$$H(X:Y) \equiv H(X) + H(Y) - H(X,Y) \quad (6.1.6)$$

Looking at the previous definitions we see $H(X:Y)$ can be recast as:

$$H(X:Y) = H(X) - H(X,Y) \quad (6.1.7)$$

6.1.4 Markov chains and data processing

Markov chains can be useful in describing random processes with many steps, for instance, the random losses of infromation during data processing. A Markov chain is a sequence of random variables $X_1 \to X_2 \to X_3 \to \cdots$ each of which is conditional on only the prior (X_{n-1}) variable.

The **data processing inequality** is the following theorem:
Supose $X \to Y \to Z$ is a Markov chain. Then

$$H(X) \geq H(X:Y) \geq H(X:Z) \tag{6.1.8}$$

The first inequality becomes an equality if an only if, given Y it is possible to reconstruct X

6.2 Shannon's 2 Theorems

6.2.1 Shannon's noiseless coding theorem

Here we state Shannon's two foundational theorems of classical information theory, published in 1948.

I Noiseless Coding Theorem

N i.i.d. (independent and independently distributed) random variables each with entropy $H(X)$ can be compressed into more than $NH(X)$ bits with negligible risk of information loss, as N tends to infinity; but conversely, if they are compressed into fewer than $NH(X)$ bits it is virtually certain that information will be lost.

Equivalent statement:
Suppose $\{X\}$ is an infromation source whose "letters" (bits) are i.i.d. Suppose the rate R is greater than the entropy rate $H(X)$. Then there exists a reliable compression scheme of rate R for the source. Conversely, if $RH(X)$ then any compression scheme will not be reliable.

II Noisy Channel Coding Theorem

6.3 Fisher Information

Fisher information is a quantity in mathematical statistics that may be very important in future physics, especially when considering the information that can be obtained from a given measurement or set of measurements. As will be discussed below, at least one physicist, B. Roy Frieden, believes that some fundamental physics can be explained, and perhaps even derived completely from the Fisher information.[2]. Fisher information was introduced 1922, whereas Shannon information was introduced in 1948.

Fisher information is a formal way of quantifying the amount of information that is contained in a set of a measurements. Given a quantity θ that we are trying to measure, there will always be some "noise" or deviations in the measurements. Repeated measurements will assume a probability density function $f(y;\theta)$, which is characterized by a random variable y whose distribution depends on θ. $y_i = \theta + x_i$, where x_i are noise factors. The estimate of the true value of θ from the data is given by an "estimator", $\hat{\theta}(y)$ which could be the mean, for example, or something more sophisticated. The derivative with respect to θ of the log of $f(y;\theta)$ is a quantity called the "score". The Fischer information is the 2nd moment of the score:

$$I_F \equiv \int_{-\infty}^{\infty} \left(\frac{\partial \ln(f)}{\partial \theta} \right)^2 f dy \tag{6.3.1}$$

The Fisher information satisfies the following inequality with the mean squared error e^2 (which is also the variance of f):

$$e^2 I_F \geq 1 \tag{6.3.2}$$

This relation is known as the Cramer-Rao inequality or Cramer-Rao bound, which gives the interpretation of the Fisher information mentioned above – as the error decreases, the Fisher information increases. In other words, measurements that are more accurate have more Fisher information.

The proof of the Cramer-Rao inequality relies of the assumption that we use an unbiased estimator $\hat{\theta}(y)$. That means,

$$\left\langle \hat{\theta}(y) - \theta \right\rangle = \int [\hat{\theta}(y) - \theta] f(y;\theta) dy = 0 \tag{6.3.3}$$

Now we take the derivative with respect to θ, yielding

$$\int [\hat{\theta} - \theta)] \frac{\partial f}{\partial \theta} dy - \int f dy = 0 \tag{6.3.4}$$

Now we use the fact that

$$\frac{\partial f}{\partial \theta} = f \frac{\partial \ln(f)}{\partial \theta} \tag{6.3.5}$$

And the fact that f is normalized, to get

$$\int [\hat{\theta} - \theta)] \frac{\partial \ln(f)}{\partial \theta} f dy = 1 \tag{6.3.6}$$

We now factor the integral as follows:

$$\int \left[(\hat{\theta} - \theta)) \sqrt{f} \right] \left[\frac{\partial \ln(f)}{\partial \theta} \sqrt{f} \right] dy = 1 \tag{6.3.7}$$

Square the equation. The Cauchy-Schwarz inequality $((\int f^2)(\int g^2) \geq [\int fg]^2)$ gives

$$\int \left[(\hat{\theta} - \theta)) \sqrt{f} \right]^2 dy \int \left[\frac{\partial \ln(f)}{\partial \theta} \sqrt{f} \right]^2 dy \geq 1 \tag{6.3.8}$$

Inspecting the terms, we see these manipulations have proven the Cramer-Roe bound:

$$e^2 I \geq 1 \tag{6.3.9}$$

6.3.1 Rényi entropy

A generalization of the Shannon entropy is the Rényi entropy, also called the Rényi information. The Rényi entropy of order α is defined as

$$H_\alpha(X) = \frac{1}{1-\alpha} \log_2 (\sum_{i=1}^{n} p_i^\alpha) \tag{6.3.10}$$

When $\alpha = 1$, this reduces to the Shannon entropy. To show this requires taking the limit of the expression using L'Hospital's rule:

$$\begin{aligned}\lim_{\alpha\to 1} H_\alpha(X) &= \lim_{\alpha\to 1} \frac{\frac{d}{d\alpha}\log_2(\sum_{i=1}^n p_i^\alpha)}{\frac{d}{d\alpha}(1-\alpha)} \\ &= \lim_{\alpha\to 1} -\frac{\sum_{i=1}^n p_i^\alpha \log_2(p_i)}{\sum_{i=1}^n p_i^\alpha} \\ &= -\sum_{i=1}^n p_i \log_2(p_i)\end{aligned} \tag{6.3.11}$$

The Rényi entropy, as with the Shannon entropy, can be used as a measure of entanglement.

6.3.2

6.4 Quantum Information Theory

6.4.1 Von Neumann Entropy

As was discussed in the section on the density matrix:

$$S(\rho) \equiv -\operatorname{tr}(\rho \ln \rho) \tag{6.4.1}$$

6.4.2 The no-cloning theorem

The no cloning theorem states that it is impossible to make a perfect copy of a quantum state. It is interesting that this fact was not discovered until 1982, when a proof was published by Wootters, Zurek and Dieks. We can prove the theorem as follows: suppose we have a "copying machine" which has two slots, A and B. We will try to use it to copy the state $|\psi\rangle$ in A to the state in B. Without loss of generality, we assume the state in slot B is a pure state $|s\rangle$. Some unitary transformation effects the copying procedure:

$$|\psi\rangle|s\rangle \to U(|\psi\rangle|s\rangle) = |\psi\rangle|\psi\rangle \tag{6.4.2}$$

Suppose that the copier is used to copy two different systems:

$$\begin{aligned}U(|\psi\rangle|s\rangle) &= |\psi\rangle|\psi\rangle \\ U(|\phi\rangle|s\rangle) &= |\phi\rangle|\phi\rangle\end{aligned} \tag{6.4.3}$$

Because U is unitary, it must preserve the inner product:

$$\begin{aligned}(\langle\psi|\langle s|)U^\dagger U(|\phi\rangle|s\rangle &= (\langle\psi|\langle\psi|)(|\phi\rangle|\phi\rangle) \\ \langle\psi|\phi\rangle &= (\langle\psi|\phi\rangle)^2\end{aligned} \tag{6.4.4}$$

This is only possible if $\langle\psi|\phi\rangle = 0$, in which case the two states or orthogonal, or if $\langle\psi|\phi\rangle = 1$, in which case $|\psi\rangle = |\phi\rangle$.

Thus we have shown that copiers are limited to copying sets of states which are orthogonal, which is, in fact, a set of measure zero in a continuous space of states. Any realistic quantum state cannot be copied pefectly. In the words of DJ Griffiths, "It's as

though you bought a Xerox machine that copies verical lines perfectly, and also horizontal lines, but completely distorts diagonals."

Incidently, if this was not the case, it would be possible to transmit information faster than light using an EPR setup.[3] Suppose we would like to transmit "1" or "0". If Alice wants to transmit 1 to Bob, she measures S_z of her particle. Bob then makes a thousand copies of his particle and measures S_z on all of them. If they are all up or down, he knows Alice measured S_z and that the message is "1". If roughly half are up and half are down, he knows Alice didn't make such a measurement and the message is "0".

6.5 Connection to thermodynamics

There is an intimate connection between information theory and thermodynamics.

6.5.1 Landauers principle

> **Landauer's Principle**: Suppose a computer erases one bit of information. Then at least $k_B T \ln(2)$ Joules of heat must be lost to the environment. (Equivalently, $k_B \ln(2)$ nats of entropy are lost)

6.5.2 The Szilard engine

The Szilard engine is similiar to Maxwell's demon engine, which converts heat to work without any other imputs. The Szilard engine features a container with one particle in it. The "Szilard demon" inserts a wall in the center of the container, confining the particle to one half. He measures which half the particle is in, and then adiabatically moves the wall to the other side.

6.6 Distance measures for Quantum Information

This section is a summary of chapter 9 of Nielsen & Chuang.

There are three measures between sets of informations worth discussing.

6.7 The classical measure: Hamming Distance

The Hamming distance is used to measure the distance between two sets of information. To compute it, simply subtract each corresponding digit (or letter) between the two sets, and then sum the differences. For instance, the difference between 1000011 and 0000010 is 2 because they only differ in their first and last digit.

6.7.1 Distance measures between two quantum states

Fidelity

The quantum version of a fidelity is a strange beast. The fidelity of two quantum states ρ and σ is defined as :

$$F(\rho, \sigma) \equiv (\sqrt{\rho^{1/2} \sigma \rho^{1/2}}) \tag{6.7.1}$$

Chapter 7

Quantum Computation

7.1 Quantum Gates

Here is a list of important quantum gates. More detailed discussion can be found in books on quantum computation.

- Pauli-X
$$X \equiv \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \tag{7.1.1}$$
- Pauli-Y
$$Y \equiv \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix} \tag{7.1.2}$$
- Pauli-Z
$$Z \equiv \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \tag{7.1.3}$$
- Hadamard
$$H \equiv \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \tag{7.1.4}$$
- Phase
$$S \equiv \begin{bmatrix} 1 & 0 \\ 0 & i \end{bmatrix} \tag{7.1.5}$$
- T-gate (or $\pi/8$ gate)
$$T \equiv \begin{bmatrix} 1 & 0 \\ 1 & e^{i\pi/4} \end{bmatrix} \tag{7.1.6}$$

7.1.1 Multi-qubit gates

If n Hadamard gates are applied to n qubits, the result is known as the **Hadamard transform**, denoted by H^n. For instance, for 2 qubits initially in the state $|11\rangle$ the result is:

(7.1.7)

Perhaps the most used gate in quantum computation is the Hadamard gate.

7.2 Quantum teleportation

Teleportation is one of the simplier quantum algorithmns and is likely to be highly utilized in future quantum hardware. The goal of quantum teleportation is to take a one-qubit quantum state $|\psi\rangle = \alpha|1\rangle + \beta|0\rangle$ and 'teleport" it to distant location using only data passed over a classical channel. This power and weirdness of this procedure will become clear after it is described.

The paper first expounding the idea was published by Charles Bennett and coauthors in 1993. It was first confirmed experimentally in 1997 by a group in Innsbruck and has subsequently been shown to work over distances of up to 16 kilometers.

To do this, Alice and Bob each need one half of an EPR pair of quibits which they generated in the past. The procedure can be described in terms of a quantum circuit. The state imputs into the circuit contains both the EPR pair, represented by $|00\rangle + |11\rangle$ and the unknown qubit, $\alpha|1\rangle + \beta|0\rangle$:

$$\begin{aligned} |\psi_0\rangle &= \frac{1}{\sqrt{2}}\left[\alpha|1\rangle + \beta|0\rangle\right]\left[|00\rangle + |11\rangle\right] \\ &= \frac{1}{\sqrt{2}}\left[\alpha|1\rangle(|00\rangle + |11\rangle) + \beta|0\rangle(|00\rangle + |11\rangle)\right] \end{aligned} \tag{7.2.1}$$

Alice sends her qubits through a cnot gate, obtaining:

$$|\psi_1\rangle = \frac{1}{\sqrt{2}}\left[\alpha|1\rangle(|10\rangle + |01\rangle) + \beta|0\rangle(|00\rangle + |11\rangle)\right] \tag{7.2.2}$$

She then applies a Hadamard gate, obtaining:

$$|\psi_2\rangle = \frac{1}{2}\left[\alpha(|0\rangle - |1\rangle)(|10\rangle + |01\rangle) + \beta(|0\rangle + |1\rangle)(|00\rangle + |11\rangle)\right] \tag{7.2.3}$$

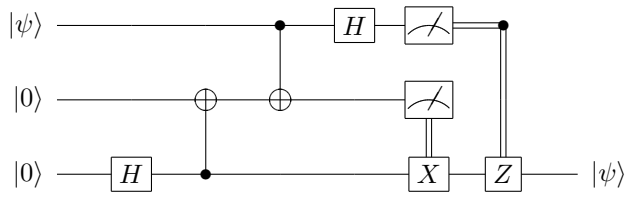
This state can be rewritten by regrouping terms. (Alices qubits are on the first two on the left now, and Bobs on the right:

$$|\psi_2\rangle = \frac{1}{2}\left[|00\rangle(\alpha|0\rangle + \beta|1\rangle) + |01\rangle(\alpha|1\rangle + \beta|0\rangle) + |10\rangle(\alpha|0\rangle - \beta|1\rangle) + |11\rangle(\alpha|1\rangle - \beta|0\rangle)\right] \tag{7.2.4}$$

Now when Alice measures the two qubits she has, we know what Bob's qubit should be, and Alice can transmit her measurements over a classical channel and then Bob can perform manipulations to bring his qubit into the state $|\psi\rangle = \alpha|1\rangle + \beta|0\rangle$. We can procede by inspection of each case, but it turns out, that we can express in compact terms what needs to be done as follows: if Alice measures $|cd\rangle$, where $c, d \in \{0, 1\}$, then Bob's qubit is in the state $X^d Z^c|\psi\rangle$. To convert his state back to the orginal $|\psi\rangle$ he needs to perform the operation $X^d Z^c$ on the qubit (remember, the Pauli

matrices square to unity, so this operation will restore the state). After this, he can be certain he has the orginal state $|\psi\rangle$ in his hands.

The circuit for quantum teleportation (including the preperation of the Bell state) is given as follows:



7.3 "Exchange of resources" and Superdense Coding

Quantum teleportation showed that the information contained in one qubit can be "exchanged" (transfered by) one EPR pair and two classical bits. This is something profound, because it gives us a "conversion" between quantum information and classical information. It also shows why information cannot be transfered faster than light here. While (at least on paper) the state $|\psi\rangle$ is transfered instananeously, no information is transfered instaneously, because of the universe's dictation that we also transfer two bits of classical infromation over a classical channel. This requirement prevents us from transfering information faster than light in this case.

In a similiar way that teleportation shows that one qubit is equivalent to one EPR qubit and two classical bits, superdense coding shows that two classical bits can be stored in one EPR qubit.

7.4 Deutsch's Algorithm

7.4.1 The DeutschJozsa algorithm

7.5 Grover's Algorithm

Classically, the fastest way to search an unstructured database is simply to check each entry one at a time, a process requiring $\mathcal{O}(N)$ operations. Grover's quantum search algorithm completes the task in $\mathcal{O}(\sqrt{N})$ operations, a polynomial-order speedup. It was later proved that this speedup is ideal. This polynomial-order ("quadratic") speedup is not as impressive as the exponential speedup provided by Shor's algorithm, but because searching is such a ubiquitous task in computer science, this algorithm has very exciting practical prospects. As we will discuss later, it can be used to speed up a large class of NP problems. The algorithm was devised in 1996 by L.K. Grover at Lucent Technologies and experimentally demonstrated in 1998 by Chuang, Gershenfeld and Kubinec.

7.5.1 The Oracle

For simplicity, we also assume that all of the items in our database are indexed by a set of states $|x\rangle$. Grover's algorithm requires a **quantum oracle** which can recognize solutions to the search problem. In order to keep the search algorithm general, we will not specify the innerworkings of this oracle and consider it as a "blackbox". The oracle contains a function f which returns $f(x) = 1$ if $|x\rangle$ is a solution to the search problem and $f(x) = 0$ otherwise. The oracle is a unitary operator O which opeates on two quibits, the index qubit $|x\rangle$ and the "oracle qubit" $|q\rangle$:

$$|x\rangle|q\rangle \xrightarrow{O} |x\rangle|q \oplus f(x)\rangle \tag{7.5.1}$$

As usual, $\oplus$ denotes addition modulo 2. The operation flips the oracle qubit if $f(x) = 1$ and leaves it alone otherwise.

In Grover's algorithm we want to flip the sign of the state $|x\rangle$ if it labels a solution. This is achived by setting the oracle qubit in the state $(|0\rangle - |1\rangle)/\sqrt{2}$, which is flipped to $(|1\rangle - |0\rangle)/\sqrt{2}$ if $|x\rangle$ is a solution:

$$|x\rangle \left(\frac{|0\rangle - |1\rangle}{\sqrt{2}}\right) \xrightarrow{O} (-1)^{f(x)}|x\rangle \left(\frac{|0\rangle - |1\rangle}{\sqrt{2}}\right) \tag{7.5.2}$$

We regard $|x\rangle$ as flipped, thus the oracle qubit is not changed, so by convention the oracle qubits are usually not mentioned in the specification of Grover's algorithm. Thus the operation of the oracle is written as:

$$|x\rangle \xrightarrow{O} (-1)^{f(x)}|x\rangle \tag{7.5.3}$$

For simplicity, we designate the correct solution to the search project (the "target") as $|t\rangle$. Then the operation of the oracle can be represented by the operator $I - 2|t\rangle\langle t|$.

If there are N items in our database, we need $n = \log_2(N)$ qubits to index the database.

7.5.2 The Algorithm

1. Use n Hadamard gates (a Hadamard transform) to initilize the state of the system to:
$$|\psi\rangle = \frac{1}{\sqrt{N}} \sum_{x=0}^{N-1} |x\rangle \tag{7.5.4}$$
2. Apply the "Grover iteration" $r(N)$ times:
   1. Apply the oracle, $I - 2|t\rangle\langle t|$
   2. Apply the Grover diffusion operator $G = -(I - 2|\psi\rangle\langle\psi|)$
3. Perform a measurement

7.5.3 Geometric visualization

Meditation on this algorithm reveals a beautiful geometric picture of how it works. This picture is most easily visualized by looking at diagram on the right.

Suppose we have N elements in our database and M search solutions. We define the non-solution vector $|\alpha\rangle$ and the the solution vector $|\beta\rangle$ as:

$$\begin{aligned} |\alpha\rangle &\equiv \frac{1}{\sqrt{N-M}} \sum_{x \in \text{non-solns}} |x\rangle \\ |\beta\rangle &\equiv \frac{1}{\sqrt{M}} \sum_{x \in \text{solns}} |x\rangle \end{aligned} \tag{7.5.5}$$

The initial state $|\psi\rangle$ is then:

$$|\psi\rangle = \sqrt{\frac{N-M}{N}}|\alpha\rangle + \sqrt{\frac{M}{N}}|\beta\rangle \tag{7.5.6}$$

The angle $\theta/2$ between $|\psi\rangle$ and $|\alpha\rangle$ is given by

$$\cos\frac{\theta}{2} = \sqrt{\frac{N-M}{N}} \tag{7.5.7}$$

The probability of sucess of the algorithm is given by the exact expression:

$$\sin^2\left[\left(r+\frac{1}{2}\right)(\pi - 2\theta)\right] \tag{7.5.8}$$

The number of iterations before the best probability is achieved is given by $r(N)$

$$r(N) = \left\lceil \frac{\pi}{4}\sqrt{\frac{N}{M}} \right\rceil \tag{7.5.9}$$

7.5.4 The Partial Search modification of Grover's Algorithm

A modifcation of Grover's algorthm is called **Partial Search**, which was first described by Grover and Radhakrishnan.[**?**] In parial search, one is not interested in find the exact address of the target item, only the first few digits of the address. Equivalently, we can think of "chunking" our search space into blocks, and then asking "in which block is the target item?". In many applications, such a search yields enough information if the target address contains the information wanted. For instance, to use the example given by L.K. Grover, if one has a list of students organized by class rank, we may only be interested in whether a student is in the lower 25%, 25-50%, 50-70% or 75-100% precentile. An example given by Korepin is that we may be just interested in what state the town of Stony Brook is, not its precise location.

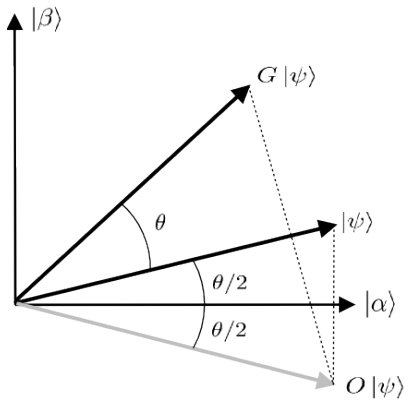
To describe partial search, we consider a database seperated into K blocks, each of size $b = \frac{N}{K}$. Obvioulsy, the partial search problem is easier. Consider the approach we would take classically - we pick one block at random, and then perform a normal search through the rest of the blocks (in set theory language, the compliment). If we don't find the target, then the we know it's in the block we didn't search. The average number of iterations drops from $\frac{N}{2}$ to $\frac{N-b}{2}$.

Grover's algorithm requires $\frac{\pi}{4}\sqrt{N}$ iterations. Partial search will be faster by a numerical factor which depends the number of blocks K. Partial search uses n_1 global interations and n_2 local iterations. The global Grover operator is designated G_1 and the local Grover operator is designated G_2.

The global grover operator acts acts on the blocks. Essentially, it is given as follows:

1. Perform j_1 standard Grover iterations on the entire database.
2. Perform j_2 local Grover iterations. A local Grover iteration is a direct sum of Grover iterations over each block.
3. Perfore one standard Grover iteration

One might also wonder what happens if one applies successive partial searches at different levels of "resolution". This idea was studying in detail by Korepin and Xu, who called it "Binary (quantum) search". They proved that it is not in fact any faster than a single partial search.



Bibliography

[1] *Path Integrals and Quantum Mechanics* By Richard Feynman & Albert Hibbs. Emended by Daniel Styer. Dover Publications, 2010.

[2] *Physics from Fisher Information: A Unification.* By B. Roy Frieden, Cambridge University Press, New York, 1998

[3] D.J. Griffiths. *Introduction to Quantum Mechanics: 2nd Edition.* 2005 Pearson Education.

[4] L.K. Grover and J. Radhakrishnan, quant-ph/0407122

[5] L.D. Landau and E.M. Lifshitz. *Vol. 3 Quantum Mechanics - Non-Relatavistic Theory Volume 3 of the Course of Theoretical Physics.* Pergamon Press

[6] M. Nakahara *Geometry, Toplogy and Physics* Taylor & Francis Group press 2003.

[7] M.A. Nielsen and I.L. Chuang. *Quantum Computation and Quantum Information* Cambridge University Press 2009.

[8] Preskill *Lecture Notes on Quantum Computation*

[9] JJ Sakurai and J. Napolitano. *Modern Quantum Mechanics* 2nd Edition Pearson 2011.

[10] R Shankar *Principles of Quantum Mechanics* 2nd Edition Springer Science 1994.

[11] *Superdeterminism* Wikipedia.