Physics 622 Spring 2000 - Dr. D. Fivel

Chapter 1

The Basic Ideas

If the criterion of a successful physical theory is the accuracy of its predictions, quantum mechanics has to be reckoned the most successful theory ever contrived. On the other hand quantum mechanics has serious conceptual difficulties such as the "measurement problem" — known also as the "Schrödinger Cat" problem — and the Einstein-Podolsky-Rosen problem (EPR). Many successful working physicists go through life ignoring these problems, following a philosophy called SUAC which means "sweep under a carpet" or "shut up and calculate". On the other hand dilligent attention to these difficulties has led in recent years to the discovery of the ideas of quantum cryptography and quantum computing. Since it is both more intellectually satisfying and probably useful for students of physics to be aware of these skeletons in the closet, they will be discussed in this course.

1.1 The Simplest Experiments Exhibiting Non-Classical Behavior.

Let us begin by examining a set of experiments with polarized light that compel us to give up the classical description and lead us to the quantum theory. In these experiments unpolarized light is passed through a filter called a "polarizer" which is labelled y and subsequently through a second filter called an "analyzer" which is labelled x. If I is the intensity after passing the y-filter, and I' the intensity after passing the x-filter, our experiments will measure

$$p(x, y) = I'/I. \tag{1.1}$$

For linearly polarized light you may recall from elementary physics that $p(x, y)$ obeys "Malus' Law" :

$$p(x, y) = \cos^2(\Theta(x, y)), \tag{1.2}$$

where $\Theta(x, y)$ is the acute angle between the polarizing axes of x and y.

Let us see how this can be deduced from the classical Maxwell theory together with its generalization to elliptic polarization:

a. Maxwell Theory of Polarization

Consider the behavior of a monochromatic, electromagnetic wave of frequency ω as it travels in the x_3 direction through a slab of *anisotropic*, dielectric material of thickness δ. We suppose there to be a distinguished direction in the plane of the slab, say the x_2 direction, for which the dielectric constant ϵ_2 and the conductivity σ_2 differ significantly from their vacuum values, i.e. from $\epsilon = 1$, $\sigma = 0$. For this discussion we use units in which the speed of light in vacuum is unity.

One will find from Maxwell's equations that the 2-component of the **E**-field is attenuated by the factor $\exp(-\mathfrak{Im}(z))$, and the phase of the 2-component is advanced relative to that of the 1-component by $\mathfrak{Re}(z)$, where

$$z = \omega\delta\left(\sqrt{\epsilon_2} + i\frac{4\pi\sigma_2}{\omega}\right), \tag{1.3}$$

$\surd$ **Exercise 1**: Show this.

A *linear polarizer* for the direction $\hat{\mathbf{u}}$ is constructed by rotating the slab so that the 2–direction is the direction perpendicular to $\hat{\mathbf{u}}$ and making $\mathfrak{Im}(z)$ so large (e.g. by increasing δ) that this component of the **E**-field is *quenched*. The light thus produced is said to be *linearly polarized in the* $\hat{\mathbf{u}}$ *direction* . We next construct a *phase shifter* by passing the linearly polarized light through a second slab for which $\mathfrak{Re}(z) \neq 0$, $\mathfrak{Im}(z) = 0$. If E_1 and E_2 are the components of the linearly polarized **E** field produced by the first slab relative to the 1–2 axes of the second slab, the phase of E_2 will be advanced by $\phi = \mathfrak{Re}(z)$ relative to the phase of E_1.

If the light emerging from this combination of slabs travels in the x_3 direction through vacuum, the **E**-field in the transverse plane will be given by:

$$\mathbf{E} = E_0 \begin{pmatrix} u_1 \cos(\omega(x_3 - t)) \\ u_2 \cos(\omega(x_3 - t) + \phi) \end{pmatrix}. \tag{1.4}$$

One readily verifies that for any fixed x_3, the vector **E** traces out an ellipse, and we say that the light is in a *state of elliptic polarization.*

$\surd$ **Exercise 2**: Show that $\hat{\mathbf{u}}, \phi$ determine the relative size of the major and minor axes of the ellipse, the angle of tilt of the major axis, and the direction of rotation around the ellipse. What are the conditions for left and right circular polarization?

One should note that reversing the direction of $\hat{\mathbf{u}}$ has no effect on the polarization state. One should also note that the effect of changing the *relative* sign of the components u_1 and u_2 of $\hat{\mathbf{u}}$ can also be achieved by $\phi \to \phi + \pi$. Thus, to obtain a one-one correspondence with the possible states of elliptic polarization, we allow ϕ to have the range $0 \leq \phi < 2\pi$ but restrict $\hat{\mathbf{u}}$ to the first quadrant or equivalently set

$$u_1 = \cos\frac{\theta}{2}, \quad u_2 = \sin\frac{\theta}{2}, \qquad 0 \leq \theta \leq \pi. \tag{1.5}$$

(The reason for using the half-angle will become clear below.) We thus denote the state by the pair of angles $y = (\theta, \phi)$ and refer to the pair of slabs that produce the state as a y-polarizer.

We construct an *analyzer* for the the state $y = (\theta, \phi)$ by first applying the relative phase change $-\phi$ to a beam and then passing the resulting beam through the linear polarizer associated with θ. A beam that has passed a y-polarizer will be called a y-beam.

$\surd$ **Exercise 3**: (a) Show that a y-beam is unattenuated by a y-analyzer. (b) If a w-beam is quenched by a y-analyzer, what is w?

Now let us consider the general case. We wish to compute the *attenuation factor* $p(y_a, y_b)$ by which the intensity of a y_a-beam is reduced when passed through a y_b-analyzer. Suppose, after the relative phase shift $-\phi_b$ produced by the analyzer, the beam arrives at time t at the $\hat{\mathbf{u}}_b$ linear polarizer slab which we take to be on the plane x_3. The $\mathbf{E}$ field will then be given by

$$\mathbf{E} = E_0 \begin{pmatrix} u_{a1} \cos(\omega(x_3 - t)) \\ u_{a2} \cos(\omega(x_3 - t) + \phi_a - \phi_b) \end{pmatrix}. \tag{1.6}$$

As we noted above, only the component in the direction of $\hat{\mathbf{u}}_b$ will pass the analyzer. Thus the amplitude of the field that passes the analyzer will be:

$$\mathbf{E} \cdot \hat{\mathbf{u}}_b = E_0 \left(u_{a1} u_{b1} \cos(\omega(x_3 - t)) + u_{a2} u_{b2} \cos(\omega(x_3 - t) + \phi_a - \phi_b) \right). \tag{1.7}$$

The intensity is proportional to the square of the electric field. Averaging $(\mathbf{E} \cdot \hat{\mathbf{u}}_b)^2$ over t and dividing by the average $(\frac{1}{2}E_0^2)$ of the beam intensity before passing the analyzer, we see that the relative intensity after passing to before passing is:

$$(u_{a1} u_{b1})^2 + (u_{a2} u_{b2})^2 + 2 u_{a1} u_{b1} u_{a2} u_{b2} \cos(\phi_a - \phi_b). \tag{1.8}$$

This can be put into the following convenient form: With each polarization state $y = (\theta, \phi)$ associate a two component, complex unit ray by

$$y \to |y\rangle = e^{i\Lambda} \begin{pmatrix} u_1 \\ u_2 e^{i\phi} \end{pmatrix} = e^{i\Lambda} \begin{pmatrix} \cos\frac{\theta}{2} \\ e^{i\phi} \sin\frac{\theta}{2} \end{pmatrix}, \tag{1.9}$$

in which $e^{i\Lambda}$ is an arbitrary unimodular phase factor.

Now from the definition of the complex scalar product given in Chapter 0, we find that $|\langle y_a | y_b \rangle|^2$ coincides with the expression (1.8) above.

$\surd$ **Exercise 4**: Show this.

Referring to Chapter 0 (eq 56) we can state our result as the following generalization of Malus' Law:

Malus' Law for Elliptic Polarization: (1) The states x of elliptic polarization are in one-one correspondence with the projectors $\boldsymbol{\pi}(()x) = |x\rangle\langle x|$, where $|x\rangle$ is any representative of a unit ray in complex 2-dimensional vector space. Equivalently the states x are in one-one correspondence with the points of the Poincaré sphere. (2) If light of intensity I_0 in the state x is incident on an analyzer for the state y, the the emerging light is in the state y and has intensity $I = I_0\, p(x, y)$ with

$$p(x, y) = \mathrm{Tr}(\boldsymbol{\pi}(x)\boldsymbol{\pi}(y)) = |\langle x | y \rangle|^2 = \cos^2\left(\frac{\Psi(x, y)}{2}\right), \tag{1.10}$$

in which Ψ is the great circle arc joining the representation points on the Poincaré sphere.

$\surd$ **Exercise 5**: (a) Where on the sphere are the states of linear polarization? (b) Where on the sphere are the states of right and left circular polarization $|\pm\rangle$. (c) If we map the sphere so as to put $|\pm\rangle$ at the poles † where will the states of linear polarization be?

b. Failure of Classical Theory

The decisive improvement in technology that led to the discovery of the failure of the classical theory of light was the invention of sensitive photocells that could respond to very low intensities. One then found that the energy was quantized in multiples of the frequency. These quanta are called *photons*.

Photons that emerge from a y-filter will be called y-photons. We then find that *it is impossible to predict whether any given y-photon will pass an x-filter.*†† However we find that *the probability $p(x, y)$ that a y-photon will pass an x-filter is given by the generalized Malus' Law above.* It is the *statistical randomness* in the behvior of individual photons which appears to be non-classical.

There is of course nothing particularly strange about having statistical fluctuations in physics. Indeed classical statistical mechanics is used to deal with systems in which there are variables over which we have no control and over which we must therefore average. If the photon transmission statistics arose in this way it would mean that there would be an uncontrolled "hidden variable" which actually *determines* whether an individual photon will get through any particular type of filter.

Before 1964 it was a commonly held view that a hidden variable explanation of quantum phenomena would be discovered someday. Indeed experiments were performed to test theories of the following type:

Let the hidden variable λ range over the interval $[0, 1]$ and suppose the rule of passage is this: A photon having passed a y-polarizer will or will not pass an x-analyzer according as λ is or is not in $[0, p(x, y)]$ with $p(x, y)$ given by (1.1). If we have no control of λ, it is equally likely to be found anywhere in $[0, 1]$ which means that it will be in $[0, p]$ a fraction p of the time. We thus recover Malus' Law if $p < 1/2$. If $p > 1/2$ we simply interchange x and x' where x' is the orthogonal filter and proceed as before.

To test a model of this kind we observe that just after a photon passes an y-polarizer and a subsequent x-analyzer, the value of λ will be in $[0, p(x, y)]$. It will thus no longer be equally likely to be anywhere in $[0, 1]$. Hence, unless it quickly *relaxes* we shall get a *deviation*

† In the optics literature it is customary to define the sphere so that the states $|\pm\rangle$ are at the poles. One may also consider a family of spheres with the radius proportional to the intensity. The cartesian components on this sphere are called "Stokes parameters". See e.g. R.W. Ditchburn, "Light".

†† Unless, of course, they happen to be parallel or perpendicular.

from Malus' Law if the photon is passed through a subsequent z-analyzer. By pressing the z analyzer close to the x-analyzer one can make the time very short for passage from the one to the other. Such experiments have been done[†] with x nearly orthogonal to y, and a distance of less than a micron between the x and z analyzers. No deviations from Malus' Law have been observed!

√ **Exercise 6**: (a) Why does one want to choose x nearly orthogonal to y? (b) Show that the absence of deviation from Malus' Law implies that the relaxation time of the hidden variable must be less than $10^{-14}s$. (c) Suppose the hidden variable didn't relax at all in the passage from x to z and suppose that $z = y$. What would we observe?

In the above experiment a photon passes one filter and then another, so that it is possible for the hidden variable to "know" what the photon already did (i.e. that it passed the y-filter) before "telling" it whether to pass the x-filter. Thus we had the freedom to consider a rule of passage employing *any* function of x and y. However, as Einstein, Podolsky, and Rosen (EPR) pointed out in their 1935 paper, there is a class of experiments in which one does not have that freedom. Rather than examining the gedanken experiment considered by EPR let us look at one that actually has been done at the University of Maryland and in Paris: [*]

The experiment makes use of a processs known as *parametric down-conversion* whereby a single photon in a non-linear optical medium is converted into a pair of photons each with half the frequency and with orthogonal polarizations. Thus while we cannot predict what the polarization of a member of the pair will be, we can predict with certainty that if member-1 has polarization x, then member-2 will have the orthogonal polarization x' and vice versa.

We now count the number of times member-1 passes an x-filter *and* its partner passes a y' filter. This is done with an arrangement such that the detection of member-2's passage cannot in any way influence member-1's passage provided that *no signals can travel faster than light*. This is accomplished by using *delayed choice* of the filter orientations. On the other hand we can be sure that member-1 *would have* passed a y-filter had we tested it.[††] *It is found that $p(x, y)$ is still given by Malus' Law when the experiment is done in this non-invasive manner.*

It follows that the hidden variable λ must independently "tell" both members what to do. Thus if $\Lambda(x)$ is the subset for which member-1 passes an x-filter, the perfect correlation means that (up to a set of measure zero), $\Lambda(x')$ must be the subset for which member-2 passes an x filter. The subset for which member-1 passes an x filter and member-2 passes a y' filter will then be $\Lambda(x) \cap \Lambda(y)$. Hence the conditional probability that member-1 pass

---

[†] C. Papaliolious, Phys. Rev. Let. **18**, 15 (1967)

[*] Y.H. Shih and C.O. Alley, Phys. Rev. Let. **61**, 2921 (1988), and A. Aspect, J. Dalibard, and G. Roger, Phys. Rev. Let. **49**, 1804 (1982).

[††] Some philosophers object to this kind of argument which is known as "counter-factual" reasoning.

an x-filter given that member-2 passes a y' filter is

$$p(x,y) = \frac{\mu(\Lambda(x) \cap \Lambda(y))}{\mu(\Lambda(y))}, \tag{1.12}$$

where μ is the probability measure on the sets.

√ **Exercise 7**: Since from (1.2) we have

$$p(x,y) = p(y,x),\ p(x,x) = 1,\ \forall x,y, \tag{1.13}$$

show that we can redefine μ so that (1.12) becomes

$$p(x,y) = \mu(\Lambda(x) \cap \Lambda(y)),\ \ \mu(\Lambda(x)) = 1,\ \forall x,y. \tag{1.14}$$

The question then is whether we can fit formula (1.2) for $p(x,y)$ to (1.14), i.e. can we find a choice of sets $\Lambda(x)$ and measure μ that reproduces the data? In Bell's 1964 paper he showed that the answer is *no*. The following is a different proof due to D. Fivel which has the advantage of revealing the essentially geometric character of the incompatibility between quantum mechanics and hidden variable theories:

Suppose we collect the measured values of $p(x,y)$ for all choices of x,y in a table which we call a p-table. We can make the p-table look like a table of road distances between cities in the following way: We consider two labels x,y identical if $p(x,z) = p(y,z)$ for all z. We then define a "distance" $d(x,y)$ between labels by

$$d(x,y) = \sup_z |p(x,z) - p(y,z)|. \tag{1.15}$$

√ **Exercise 8**: Using only (1.15) and that p is a real number show that $d(x,y)$ enjoys the following properties:

$$d(x,y) = 0 \Longleftrightarrow x = y,\ \ d(x,y) = d(y,x),$$

$$d(x,y) + d(y,z) \geq d(x,z),\ \forall x,y,z \quad (\text{ triangle inequality }). \tag{1.16}$$

Thus d is a *metric* on the p-table. Now if (1.14) is used in computing $d(x,y)$ from (1.15) one observes that

$$d(x,y) = 1 - p(x,y). \tag{1.17}$$

√ **Exercise 9**: Prove this. Hint: Note that overlaps between sets cancel because of the minus sign on the right side of (1.15). When the triangle inequality in (1.16) is written in terms of p we obtain

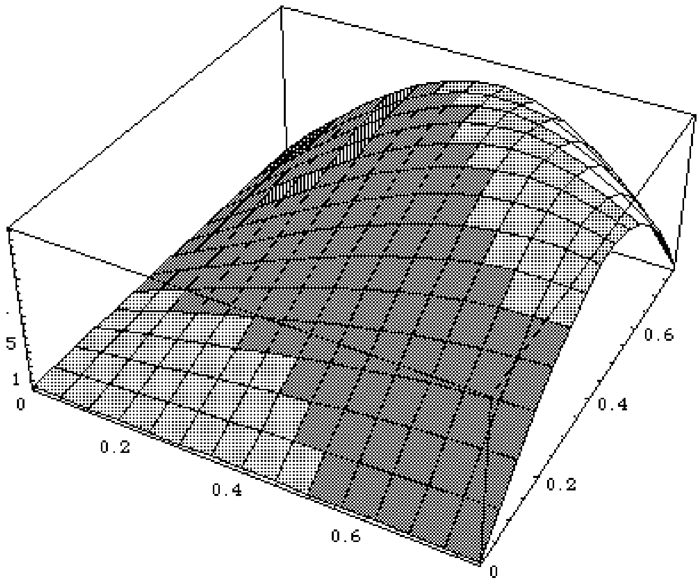
$$p(x,y) + p(y,z) \leq 1 + p(x,z),\ \ \forall x,y,z, \tag{1.18}$$

which is known as *Bell's inequality*.

Thus hidden variables can be ruled out if (1.2) violates this inequality. Suppose polarizers x,y,z have axes at angles $0, \psi, \psi + \phi$ respectively. Then (1.18) gives

$$R(\psi, \phi) \equiv \frac{\cos^2 \psi + \cos^2 \phi}{1 + \cos^2(\psi + \phi)} \leq 1. \tag{1.19}$$

The left side is plotted below for $0 \leq \phi, \psi \leq \pi/4$ and is seen to violate Bell's inequality.



We will see later that the correct quantum mechanical expression for $p(x, y)$ replaces (1.17) with

$$d(x, y) = \sqrt{1 - p(x, y)}. \tag{1.20}$$

For the three linear polarizers above the triangle inequality now becomes

$$\sin \psi + \sin \phi \geq \sin(\psi + \phi). \tag{1.21}$$

√ **Exercise 10**: Show that the triangle inequality is always satisfied by relating the sines to the lengths of chords on a circle.

Einstein died before Bell's paper appeared and would, no doubt, have been very disturbed by it. The 1935 EPR paper referred to above gave a strong argument that without hidden variables quantum mechanics was "incomplete".† They first defined a "realistic" theory as one in which any property that we can predict without in any way disturbing a system is one that belongs to the system. They then defined a "complete" theory as one in which all properties that belong to a system are accounted for. Next they constructed a gedanken experiment in which (if it could be done) there would appear *correlations between space-like*

† The term EPR "paradox" is a misonomer. A paradox is a logical inconsistency. What EPR did was to prove a theorem that is *puzzling* but not paradoxical.

separated systems as there does in the down-conversion experiment described above. But space-like separated correlations imply the possibility of prediction without disturbance if nothing super-luminal happens. Hence realism requires that there be some property of the systems that carries the correlation, and since quantum mechanics doesn't account for it, quantum mechanics is incomplete.

The most common view these days is that quantum mechanics is "harmlessly" non-local. To understand what that means consider a random sequence of twenty-five outcomes for member-1: 0 indicates that it does not get through a certain type of filter and 1 indicates that it does.

$$0, 0, 1, 1, 0, 1, 0, 0, 0, 1, 0, 0, 0, 1, 1, 0, 1, 1, 1, 0, 0, 0, 1, 0, 0$$

When we inspect the corresponding list for member-2 each of which was recorded *simultaneously* but far away from member-1 we find:

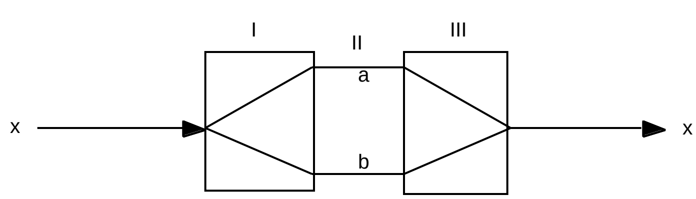
$$1, 1, 0, 0, 1, 0, 1, 1, 1, 0, 1, 1, 1, 0, 0, 1, 0, 0, 0, 1, 1, 1, 0, 1, 1,$$

The two lists are seen to be perfectly correlated. *If we could control the outcome for member-1 we could use this to send messages to an observer at the location of member-2. But since we cannot control the outcome there is no way to take advantage of the correlation to send messages.*

Thus there is a kind of "peaceful coexistence" between quantum mechanics and relativity, but the situation is still considered very spooky by everyone who has studied it even now some sixty-five years after the EPR paper.

1.2 Superpositions and Mixtures - The Measurement Problem

In certain crystalline materials such as feldspar the dielectric constant ϵ and/or permeability μ are different for light travelling in different directions, so that these quantities must be described mathematically by tensors rather than scalars. This results in the phenomenon of *birefringence* wherein an incident electromagnetic wave is split into two waves that follow different paths and have different polarizations. It is possible to put two such crystals together as indicated schematically in the figure below, so that the two subbeams are recombined into a single beam.† Let light in polarization state x be incident on the crystal.



† This experiment is essentially the same as the "two slit" experiment in which spatial interference is detected instead of polarization interference.

Suppose that according to classical (Maxwell) theory an incident wave Ψ_x will evolve into a superposition $\Psi_a + \Psi_b$ with Ψ_a, Ψ_b localized in the upper and lower parts of region II. The ratio of intensities in these two regions will be $|\Psi_a|^2/|\Psi_b|^2$, and if we reduce the intensity so that one photon at a time is in the apparatus, this will be the ratio of the probabilities for finding it in these two regions. The emergent beam is found to be in the same state x as the incident state unless we try to determine which path individual photons follow. Such a determination requires interacting with the photons which randomizes the relative phases and destroys the interference.

The quantum theory expresses this behavior as follows. The dynamics determines an evolution of the incident state $|x\rangle$ into a linear combination (superposition) of states $|a\rangle, |b\rangle$ of the form

$$|z\rangle \to \alpha|a\rangle + \beta|b\rangle, \quad |\alpha|^2 + |\beta|^2 = 1. \tag{1.22}$$

Thus the projector $\boldsymbol{\pi}(z)$ evolves according to

$$\boldsymbol{\pi}(z) \to (\alpha|a\rangle + \beta|b\rangle)(\alpha^*\langle a| + \beta^*\langle b|) = |\alpha|^2|a\rangle\langle a| + |\beta|^2|b\rangle\langle b| + \alpha\beta^*|a\rangle\langle b| + \beta\alpha^*|b\rangle\langle a|. \tag{1.23}$$

If we do not disturb the beam in region II, the process is reversed in region III, and the emergent state is once again described by the projector $\boldsymbol{\pi}(z)$. If, however, we interact with the photon to determine which way it went, the *relative* phases of $|a\rangle$ and $|b\rangle$ become randomized so that the cross term averages to zero. As a result the beam in region II is converted to one described by

$$\boldsymbol{\pi}_m = |\alpha|^2\boldsymbol{\pi}(a) + |\beta|^2\boldsymbol{\pi}(b). \tag{1.24}$$

We refer to this as a *mixture*. The operator $\boldsymbol{\pi}_m$ which is a combination of projectors with non-negative coefficients that add up to unity is called a *density operator*.

The process involved in the transformation of the projector in (1.23) into the density operator (1.24) is the source of one of the fundamental problems of quantum theory that is still unsolved — the so-called *measurement problem*. Up to the point where the cross terms were eliminated, the dynamic evolution of the state is linear and deterministic. It is linear because it can be implemented by a unitary transformation and deterministic because, given the initial state, the state at later times is determined. But the phase randomization process leading to (1.24) is non-linear. Moreover the mixture $\boldsymbol{\pi}_m$ describes a stream of photons that are sometimes in the state $\boldsymbol{\pi}(a)$ and sometimes in the state $\boldsymbol{\pi}(b)$ with a distribution determined by the coefficients.

The problem thus arises because the theory is silent as to how and when linear, deterministic evolution becomes non-linear and stochastic (probabilistic). It is customary to refer to the transformation occuring in going from (1.23) to (1.24) as “collapse of the wave function”, and hence the measurement problem is often characterized as one of explaining wave function collapse.

Schrödinger himself was well-aware of the measurement problem and in fact dramatized its importance by a celebrated example known as the "Schrödinger Cat". The idea is this: Quantum mechanics deals with microscopic phenomena where, because we have no common sense experience, we are willing to live with the idea that a system can be in a superposition of states e.g. $\alpha|a\rangle + \beta|b\rangle$. However, if quantum theory is a theory of everything, then it must explain the interaction of microscopic systems with macroscopic systems. Now it may happen that if the microsystem is in state $|a\rangle$ the interaction leads to an outcome A, whereas if the microscopic system is in state $|b\rangle$ the interaction leads to a *dramatically different* outcome B. Thus if the microscopic system is in the superposition state $\alpha|a\rangle + \beta|b\rangle$ we would have a macroscopic system in a superposition of the states A and B.

In Schrödinger 's example the microscopic state $|a\rangle$ is an alpha-particle inside a nucleus and $|b\rangle$ is the state with the alpha-particle outside. In the quantum theory of radioactive decay the wave function is in a superposition of these two states. Schrödinger then imagines having an emitted alpha-particle triggering a Geiger counter whose click opens a vial of cyanide that kills a cat.† Thus when the alpha-particle is in a superposition of the inside and outside state, the combined system is in a superposition of one with a living and a dead cat. Thus we have the evolution:

$$|\text{ alpha inside }\rangle|\text{ live cat }\rangle \rightarrow$$

$$\alpha|\text{ alpha inside }\rangle|\text{ live cat}\rangle + \beta|\text{ alpha outside}\rangle|\text{dead cat }\rangle. \qquad (1.25)$$

Since we do not observe such states, its collapse to a mixture of $|\text{alpha inside }\rangle|\text{ live cat}\rangle$ with probability $|\alpha|^2$ or $|\text{ alpha outside }\rangle|\text{ dead cat}\rangle$ with probability $|\beta|^2$ must occur very rapidly. But, as we have noted, the theory gives no account of this process whatsoever.

There have been various approaches to the measurement problem. In the SUAC approach one ignores it, which is tantamount to asserting that quantum mechanics is merely a recipe for computing the behavior of a system in the face of "noise" generated by interaction with the rest of the universe. Another approach known as the "many-worlds" theory asserts that collapse doesn't take place but that the universe in some sense continually "bifurcates". Finally there are so-called dynamic reduction theories in which one attempts to construct non-linear mechanisms that bring about the collapse. The advantage of such theories over the others is that they make testable predictions.*

1.3 Properties of Polarization Mixtures

† One should not infer from this that Schrödinger was a Nazi. He actually fled Germany during WWII to avoid helping the Nazis. He went to Ireland so as not to have to help Germany's enemies either. There he spent his time in his favorite pursuit.

* I myself am the author of two such theories, one of which has made a prediction about CP violation in B meson decay which may be tested in the near future.

The density operator of a mixture contains all of the information we have about the mixture. The most general mixture of polarized light will be described by a density operator

$$\rho = \sum_j \nu_j \boldsymbol{\pi}(x_j), \quad \nu_j \geq 0, \quad \sum_j \nu_j = 1. \tag{1.26}$$

in which the states x_j are completely arbitrary. The sum can be finite or infinite.

√ **Exercise 11**: Show that (a) ρ is a positive, hermitian operator with trace equal to unity. (b) Show that all of its eigenvalues lie in the interval $[0,1]$. (c) Show that there must exist a pair of orthogonal states x, x' and and a number λ in the interval $[0,1]$ such that

$$\rho = \lambda \boldsymbol{\pi}(x) + (1-\lambda)\boldsymbol{\pi}(x'). \tag{1.27}$$

We call this the *diagonal representation of the mixture.*

Consider the four polarization states:

$$|u\rangle \text{ “}=\text{” } \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \; |d\rangle \text{ “}=\text{” } \begin{pmatrix} 0 \\ 1 \end{pmatrix},$$

$$|+\rangle \text{ “}=\text{” } \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ i \end{pmatrix}, \; |-\rangle \text{ “}=\text{” } \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -i \end{pmatrix}, \tag{1.28}$$

in which the columns are the representations of the states in the $|u\rangle, |d\rangle$ basis. With the ususal conventions the first two are orthogonal states of linear polarization and the second two are orthogonal states of circular polarization. Observe that

$$|u\rangle = \frac{1}{\sqrt{2}}|+\rangle + \frac{1}{\sqrt{2}}|-\rangle. \tag{1.29}$$

Here we have a typical interference effect in which a cancellation results in the identity of a linearly polarized state with a superposition of circularly polarized states.

The corresponding projectors are expressed by matrices in this basis, namely:

$$\boldsymbol{\pi}(u) = |u\rangle\langle u| \text{ “}=\text{” } \begin{pmatrix} 1 \\ 0 \end{pmatrix} (1,0) = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix},$$

$$\boldsymbol{\pi}(d) = |d\rangle\langle d| \text{ “}=\text{” } \begin{pmatrix} 0 \\ 0 \end{pmatrix} (1,0) = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix},$$

$$\boldsymbol{\pi}(+) = |+\rangle\langle +| \text{ “}=\text{” } \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ i \end{pmatrix} \frac{1}{\sqrt{2}} (1,-i) = \frac{1}{2} \begin{pmatrix} 1 & -i \\ i & 1 \end{pmatrix},$$

$$\boldsymbol{\pi}(-) = |-\rangle\langle -| \text{ “}=\text{” } \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -i \end{pmatrix} \frac{1}{\sqrt{2}} (1,i) = \frac{1}{2} \begin{pmatrix} 1 & i \\ i & 1 \end{pmatrix}. \tag{1.30}$$

Now observe that

$$\frac{1}{2}\boldsymbol{\pi}(u) + \frac{1}{2}\boldsymbol{\pi}(d) = \frac{1}{2}\boldsymbol{\pi}(+) + \frac{1}{2}\boldsymbol{\pi}(-) = \mathbf{I}, \tag{1.31}$$

where $\mathbf{I}$ is the *unit operator*. A mixture consisting of equal fractions of two orthogonal states of polarization is said to be *unpolarized.* What we have seen here is an example of a general theorem

Theorem: All unpolarized mixtures are *indistinguishable.*

This is quite remarkable because it says that such mixtures appear to have "forgotten" how they were formed, i.e. whether they were formed from linearly or circularly polarized states!

Notice that in computing the sum of the circularly polarized projectors there is a cancellation of the off-diagonal "i"'s in the matrix representation. The important lesson is that *interference can express itself not only in terms of the superposition of kets as in (29), but also in the indistinguishability of mixtures formed in different ways.*

Since the states of polarization can be identified with points on the Poincaré sphere, it is natural to inquire whether mixtures can be represented in a geometric way. Indeed this is the case: Let $|a\rangle, |b\rangle$ be asspcoated with points A, B on the sphere and let P be on the *chord* joining A, B. If $\nu = PA/PB$ is the ratio of the distances from P to A and P to B, we can identify the point P with the mixture $\lambda\boldsymbol{\pi}(a) + (1-\lambda)\boldsymbol{\pi}(b)$ where the ratio of λ to $(1-\lambda)$ is $1/\nu$.

This construction has some remarkable consequences: We have seen that any mixture has a diagonal representation. Now we can see that every mixture also has a "balanced" representation, i.e. as a mixture with equal proportions of two states (which may be non-orthogonal).

$\surd$ **Exercise 12**: Explain how you can use the above geometric construction to find a balanced mixture that is indistinguishable from any diagonal mixture and vice versa.

$\surd$ **Exercise 13**: A certain mixture consists of a fraction 1/4 of left-circularly polarized light, 1/4 of right circularly polarized light, and 1/2 of linearly polarized light. Write it both in diagonal and balanced form

1.4 Higher Dimensional Systems and Axioms for Quantum Mechanics

Up to this point we have discussed only the quantum mechanics of polarized light where we were led to a two-dimensional Hilbert space model. There are known systems exhibiting quantum mechanical behavior for which we are similarly led to the use of Hilbert spaces of all dimensions. However, the principles and methods developed above go over to these spaces without change. This has suggested that there ought to be an elegant and economical set of postulates from which the rules could be deduced just as one deduces the theorems of Euclidean geometry from a small set of axioms.[†]

There is of course no unique set of axioms that will do the job, so the criterion for a good set of axioms is purely aesthetic. Those with a very mathematical bent may be happy with a set of axioms that are mathematically elegant even if they are very abstract and far removed from direct contact with experiment. For example you will find textbooks with axioms like: "For every measureable quantity there is a hermitian operator."

John von Neumann and G.D. Birkhoff who were among the first "axiomatizers" for quantum mechanics constructed axioms based on the idea that quantum interference between measurements can make the use of phrases like "A and B is true" ambiguous. They then define a non-Boolean logic in which "and" is replaced by vector space intersection and obtain a small list of possible models. Essentially their axioms turn out to be equivalent to those of projective geometry.

Another approach to axiomatics has been to find the simplest set of experimental facts from which one can deduce the rules of quantum mechanics. That is, of course, in the spirit of Euclid wherein the properties assumed are those which seem most obvious in everyday experience. The advantage of such axioms is that they suggest experimental tests of the rules, i.e. they reveal experimental consequences of changes in the rules. The following example illustrates this:

We have been assuming that quantum states are to be described by vectors in a vector space over the complex numbers C. The complex numbers are commutative. Now it is possible to construct vector spaces over other kinds of numbers. In particular there is a kind of number called a "quaternion" (Q) which forms a *non-commutative* field. It is possible to construct a form of quantum mechanics over Q, indeed a whole book has been written on the subject recently.[*] There are experimental consequences of modifying quantum mechanics in this way.

In fact the situation is quite analogous to that of plane geometry. When you learn geometry in high school you formulate it in terms of the Euclid axioms which are very physical — like "two points determine a line". You then work hard to prove theorems with logic based on these axioms until you graduate to analytic geometry where you can prove theorems much more easily with Cartesian coordinates. In fact nearly everyone thinks that Cartesian and Euclidean geometry are the same thing. Actually they are not, i.e. there are theorems

† Newton's *Principia* is constructed in this way. J. M. Keynes (the economist), who made a hobby of understanding what made Newton tick, says that he figured out mechanics intuitively and then constructed all of the Euclidean style axiomatization afterwards. That is undoubtedly true.

* S.Adler, *Quaternionic Quantum Mechanics and Quantum Fields*, Oxford University Press, 1995

such as Pappus' Theorem[†] that can be deduced with Cartesian coordinates but cannot be deduced from Euclid's axioms. The reason is that there are spaces with coordinates like quaternions that obey Euclid's axioms but in which Pappus Theorem doesn't hold.

If we frame axioms for quantum mechanics in terms of directly observable experimental facts, and if we can deduce the conventional model, i.e. the Hilbert space over C model, from those facts, then anyone proposing to allow quantum mechanics over the field Q would be obligated to go to the laboratory and refute the experiments used in constructing the axioms.

I will give you an idea of what experimentally based axioms for quantum mechanics may look like. First of all the axioms must be expressed in terms of quantities that we measure, i.e. transition probabilities $p(x, y)$. Thus we can define a' as orthogonal to a by $p(a, a') = 0$. Now consider the following as potential axioms:

(A) The sum of the fractions of an x-beam that get through a y filter and a y' filter is unity.

(B) Every balanced mixture is indistinguishable from some diagonal mixture.

The second is an acceptable axiom because indistinguishability of mixtures is based on the identity of transition probabilities.

Suppose then we take (A) and (B) as axioms. Let $a, a',\ b, b',\ c, c'$ be orthogonal pairs. Assume $a \neq b'$. From (A)

$$p(a, z) + p(a', z) = p(b, z) + p(b', z) = p(c, z) + p(c', z) = 1,\ \forall z. \tag{1.32}$$

From (B) for any a, b there exists a λ in $[0, 1]$ and a c such that

$$\frac{1}{2}p(a, z) + \frac{1}{2}p(b, z) = \lambda p(c, z) + (1 - \lambda)p(c', z),\ \forall z. \tag{1.33}$$

Combining (1.32,1.33) we have:

$$p(a, z) - p(b', z) = (2\lambda - 1)(p(c, z) - p(c', z)). \tag{1.34}$$

Taking the supremum over z, and recalling the definition (1.15) we have

$$d(a, b') = |2\lambda - 1| d(c, c') = |2\lambda - 1|. \tag{1.35}$$

† On a plane draw two lines that intersect. On one of them mark off three points A, B, C and on the other mark off three points A', B', C'. Connect A to B' with a line and A' to B with a line and let P be their intersection. Similarly construct points Q from the intersection of BC' and CB' and R from the intersection of CA' with $C'A$. You will then see that P, Q and R are *colinear*.

(Why is $d(c, c') = 1$?)

Since we assume $a \neq b'$ it follows that $\lambda \neq 1/2$. From (1.32,1.33) with the choices $z = a$ and $z = b$ we get

$$p(a, b) = (2\lambda - 1)(2p(c, a) - 1) = (2\lambda - 1)(2p(c, b) - 1), \tag{1.36}$$

so that since $\lambda \neq 1/2$ there follows

$$p(c, a) = p(c, b). \tag{1.37a}$$

Hence, choosing $z = c$ in (1.33), we obtain

$$\lambda = p(c, a), \tag{1.37b}$$

so that from (1.36)

$$|2\lambda - 1| = \sqrt{p(a, b)} = \sqrt{1 - p(a, b')}. \tag{1.38}$$

Hence from (1.35) with the substitution $x = a$, $y = b'$ we have

$$d(x, y) = \sqrt{1 - p(x, y)}. \tag{1.39}$$

Now recall from Exercise 9 that local hidden variable theories have an expression for d with *no square root* on the right side of (1.39). Thus *the properties embodied in (A), (B) are enough to rule out hidden variables!*

Now let us show that quantum mechanics gives (1.39) with the square root: First we take note of a very simple but powerful theorem that will be used often.

Theorem If W is hermitian and $|z\rangle$ is any unit vector, then $|\langle z|W|z\rangle|$ assumes its extreme values when $|z\rangle$ is an eigenvector of W. The largest value is the is the largest of the absolute values of the eigenvalues of W.

$\surd$ **Exercise 14**: Prove this theorem.

Now observe that:

$$d(x, y) = \sup_z |p(x, z) - p(y, z)| = \sup_z |\langle z|\boldsymbol{\pi}(x) - \boldsymbol{\pi}(y)|z\rangle|. \tag{1.40}$$

But $W \equiv \boldsymbol{\pi}(x) - \boldsymbol{\pi}(y)$ is hermitian so by the above theorem we need only find the eigenvalue of W with the largest absolute value.

$\surd$ **Exercise 15**: Using the fact that projectors are "idempotent", i.e. $\boldsymbol{\pi}(x)^2 = \boldsymbol{\pi}(x)$ show that

$$W^3 = (1 - p(x, y))W, \tag{1.41}$$

and that this implies that the eigenvalue of W with the largest absolute value is

$\sqrt{1-p(x,y)}$.

It thus follows that:

$$d(x,y) = \sqrt{1-p(x,y)}. \tag{1.42}$$

We see then that the properties (A), (B) are good candidates for experimentally based axioms in that they eliminate hidden variables and imply the observed metric structure. It is remarkable that the properties A,B determine *almost* all of the structure of quantum mechanics. Indeed as I have shown† one can deduce that the only systems that obey axioms A, B (as generalized to more than two dimensions) are Hilbert spaces over one of the following fields: R, C, Q, Cay where R is the real field, C is the complex field, Q is the quaternion field, and Cay is the field of octonions or "Cayley numbers". For all dimensions higher than two, Cay can be eliminated. The question of what compelling additional axioms would most neatly exclude the exotic fields Q, Cay is one about which I think a great deal but do not have what I consider to be a totally satisfying answer.

1.5 Entanglement

Examine equation (1.25) which described the evolution of the Schrödinger Cat. On the right side you see what looks like a product of kets. Perhaps you noticed that I never defined such objects which are called "tensor products". Let me do so now. The tensor product of two Hilbert spaces $\mathcal{H}$ and $\mathcal{H}'$ both over C but possibly of different dimensions consists of pairs, one from the first and one from the second space. The important rule is that you can move scalars through the product and that if either factor is zero the product is zero. Scalar products are calculated in the obvious way: The scalar product of $|a_1\rangle|a_2\rangle$ with $|b_1\rangle|b_2\rangle$ is

$$(\langle a_1|\langle a_2|)(|b_1\rangle|b_2\rangle) = \langle a_1|b_1\rangle\langle a_2|b_2\rangle \qquad (1.43).$$

The rules are then extended to linear combinations of products in the obvious way.

Purists write tensor products as $|a_1\rangle \otimes |a_2\rangle$. I will do this on occasion if needed to avoid confusion or for emphasis.

Tensor product states consisting of more than one term are called "entangled" states. Their importance was recognized by Schrödinger who invented the term entanglement (German: Verschränkung).

Note that it is entanglements between microscopic and macroscopic states that must collapse rapidly to avoid the Schrödinger Cat problem. For this reason some people call entangled states $Schrödinger Kets$.

The role of entanglement in the measurement problem can be seen very clearly from von Neumann's characterization of the measurement process: Suppose H is some hermitian

† D. Fivel, Phys. Rev. A, 1993

operator with eigenstates $|n_j\rangle$ and corresponding eigenvalues λ_j, $j = 0, 1, 2,$. At time $t = 0$ a measuring device is supposed to be in a particular one of these states say $|n_o\rangle$ called "ready". When a microsystem in the state $|\phi_j\rangle$ interacts with the device in the state ready it drives the device into the state $|n_j\rangle$ by some sort of linear process indicated by:

$$|\phi_j\rangle|n_o\rangle \rightarrow |\phi_j\rangle|n_j\rangle \tag{1.44}$$

A register on the device then then records the number λ_j. Now if the microsystem is in the state

$$|\phi\rangle = \sum_j \alpha_j|\phi_j\rangle \tag{1.45}$$

when it encounters the device, linearity of the evolution implies:

$$|\phi\rangle|n_o\rangle \rightarrow \sum_j \alpha_j|\phi_j\rangle|n_j\rangle. \tag{1.46}$$

Thus the measurement process produces an entangled state, and it is this that must undergo collapse to *register* the measurement or as some say for the measurement to "happen". In the collapse the state is converted into the mixture:

$$M = \sum_j |\alpha_j|^2 \boldsymbol{\pi}(n_j). \tag{1.47}$$

The formalism now works out very elegantly: The operator H is represented by

$$H = \sum_j \lambda_j \boldsymbol{\pi}(n_j), \tag{1.48}$$

as you can verify by noting that this has the same eigenvalues and eigenvectors as H. Then

$$Tr(HM) = \sum_j \lambda_j |\alpha_j|^2 \tag{1.49}$$

which is just the expectation value of H for the mixture M. This gives us the basic rule for quantum mechanical computation:

Every measurement is associated with a hermitian operator H whose expectation value for the mixture M is $Tr(HM)$.

The concept of entanglement is also relevant to the EPR problem. You will recall that the states produced by parametric down conversion had a property of perfect correlation that was exploited in restricting the allowed form of a possible hidden variable. Let us see that it is entanglement that brings this about.

Let $\mathcal{U}$ be some fixed anti-unitary transformation and for any ket $|x\rangle$ define

$$|\widetilde{x}\rangle = \mathcal{U}|x\rangle. \tag{1.50}$$

$\surd$ **Exercise 16**: Show that for any $|x\rangle, |y\rangle$

$$\langle\widetilde{x}|\widetilde{y}\rangle = \langle y|x\rangle. \tag{1.51}$$

Now let $\mathcal{H}$ be an N dimensional Hilbert space, let $|n\rangle,\ n = 1, \cdots, N$ be a basis, and let $|S\rangle$ be the entangled, two-particle state in $\mathcal{H} \otimes \mathcal{H}$ defined by:

$$|S\rangle = \frac{1}{\sqrt{N}} \sum_{j=1}^{N} |n\rangle|\widetilde{n}\rangle. \tag{1.52}$$

Let

$$|x, y\rangle \equiv |x\rangle|y\rangle. \tag{1.53}$$

Then

$$\langle x, \widetilde{y}|S\rangle = \frac{1}{\sqrt{N}} \sum_{j=1}^{N} \langle x|n\rangle\langle\widetilde{y}|\widetilde{n}\rangle = \frac{1}{\sqrt{N}} \sum_{j=1}^{N} \langle x|n\rangle\langle n|y\rangle = \frac{\langle x|y\rangle}{\sqrt{N}}. \tag{1.54}$$

It follows from this that if a system is in the state $|S\rangle$ there is an equal probability (1/N) for finding particle-1 in any state $|x\rangle$ (i.e. if no measurement is made on particle-2). But if particle-2 is found in the state $|\widetilde{x}\rangle$ particle-1 will *always* be found in state $|x\rangle$. These are precisely the properties needed for the state to exhibit EPR correlations.

$\surd$ **Exercise 17**: Use (1.54) to verify these assertions.

States of the form (1.52) are called *generalized Bell states*.

$\surd$ **Exercise 18**: The state that occurs in the parametric down conversion experiment corresponds to $N = 2$ and the choice of $\mathcal{U}$

$$\mathcal{U}|u\rangle = |d\rangle, \quad \mathcal{U}|d\rangle = -|u\rangle. \tag{1.55}$$

Compute the state $|S\rangle$ and verify directly that there will be an EPR correlation between $|x\rangle$ and $|\widetilde{x}\rangle$ for any $|x\rangle$.