

ADELPHI UNIVERSITY

Honors College Thesis

Submitted in partial fulfillment of the requirements for
graduation from the Honors College

A Double-Slit Quantum-Eraser Experiment
Using Momentum-Entangled Photons

Justin Dove*

Advised by Prof. Sean J. Bentley[†]

May 28, 2013

Defended from:

Prof. Sean J. Bentley

Prof. John Dooher

Prof. Gregory J. Mercurio

*justindove@mail.adelphi.edu

[†]bentley@adelphi.edu

Abstract

In this paper, I discuss a proposal for and preliminary results of a double-slit quantum-eraser experiment. This experiment intends to generate double-slit interference from one half of a momentum-entangled photon ensemble by performing a local transformation on the other half, and subsequently coincidence detecting the entangled pairs. This paper is divided into two major parts. Part I is a general introduction to quantum theory. Part II discusses the details surrounding the experiment itself. In its totality, this paper should be approachable for someone with a basic physics background and experience with complex linear algebra. A reader familiar with quantum mechanics may find it an easier read, but such familiarity is unnecessary.

A Note Regarding Navigation

The digital copy of this paper makes extensive use of the `hyperref` package for L^AT_EX for the reader's convenience. Clicking the email addresses on the title page will open a new email properly pre-addressed. Clicking the number in any Part, Chapter, Section, Figure, Equation, or page reference will bring the reader immediately to the location of the referenced object. Likewise, clicking any bracketed citation number will bring the reader to the corresponding bibliography entry. Clicking any journal reference or arXiv reference in a bibliography entry will bring the reader to the web page for the article. Digital Object Identifiers (DOI's) are used where available to ensure link stability in the future. Clicking any book reference in a bibliography entry will bring the reader to its Google Books page. There the book may be previewed, searched, or purchased. URLs in bibliography entries directly link to the displayed address. The clickable regions do not extend over the authors' names. For ease of navigation, PDF bookmarks are used to create an expandable contents list that should be displayed in the left sidebar of most PDF readers. Clicking on any element of the list will bring the reader to the appropriate place in the paper. Additionally, all elements of the Contents and List of Figures can be clicked. The clickable regions do not extend over the page numbers in these lists. All of these active links may be surrounded by colored squares in some PDF readers, but such artifacts will not appear on printed copies.

Please share it with loved ones and hoard it from the non-believers.

—Burton Baton

Contents

I Background Information	7
Preface to Part I	8
1 The Postulates of Quantum Theory	10
1.1 The State Space Postulate	10
1.2 The Measurement Postulate	13
1.3 The Evolution Postulate	18
1.4 The Composite System Postulate	24
2 The Uncertainty Principle	27
2.1 The Old Uncertainty Principle	27
2.2 The Modern Uncertainty Principle	28
2.3 The Energy-Time Uncertainty Principle	31
3 Mixed States	34
3.1 Density Operators	35
3.2 Measurements	37
3.3 Evolution	38
3.4 Composite Systems	39

4 Entanglement	42
4.1 Pure Entangled States	42
4.2 General Entanglement and the Minus Sign	43
4.3 The EPR Thought Experiment	45
4.4 The Bell and CHSH Inequalities	47
4.5 Entanglement as a Resource	55
5 The Double-Slit Experiment	57
5.1 The Party Circa 1800	57
5.2 The Quantum Double-Slit Experiment	62
5.3 Complementarity	68
II The Experiment	72
Preface to Part II	73
6 Inspiration	74
6.1 Old Ideas	74
6.2 A New Idea	76
7 Previous Experiments	80
7.1 Quantum Erasers	80
7.2 The Delayed-Choice Quantum Eraser	83
7.3 How This Experiment Differs	84
8 The Apparatus	86
8.1 A Theoretical Sketch	86
8.2 Physical Realization	89

9 Results	93
9.1 Alignment Issues	93
9.2 Preliminary Results	95
10 Future Work	99
10.1 Verification of Assumptions	99
10.2 Delayed Choice	100
10.3 Incoherent Pump	102
Acknowledgments	103
Bibliography	104

List of Figures

4.1	A plot of the Bell correlations of the spin-½ singlet state.	50
5.1	A sketch of the classical double-slit experiment.	59
5.2	Photos of single-slit diffraction and double-slit interference patterns generated by laser light.	59
5.3	A rough graphical argument showing how a wave theory is sufficient to describe a double-slit interference pattern.	61
5.4	A photo of double-slit type interference of water ripples.	62
5.5	The buildup of a double-slit interference pattern from single electrons.	65
7.1	The double-slit eraser performed by Walborn et al. [82].	81
7.2	A sketch of the delayed-choice quantum eraser performed by Kim et al. [85].	83
8.1	A sketch of the title experiment apparatus.	88
8.2	A photo of the physical realization of the title experiment apparatus.	90
9.1	A data set demonstrating double-slit interference of a visible laser used in a forward alignment configuration.	96
9.2	The coincidence-count pattern minus statistically expected random counts. .	97
9.3	A rough fit of the coincidence pattern plotted on top of the alignment pattern.	98

Part I

Background Information

Preface to Part I

Part I is dedicated to introducing the basics of quantum theory to a reader familiar with complex linear algebra and includes all the relevant background information necessary to understand Part II. Readers familiar with quantum mechanics can skip to Part II. This part can be taken to stand entirely on its own as a general introduction to quantum theory. The reader is assumed to have a working knowledge of complex vector spaces, inner products, matrices, eigenvectors, and eigenvalues, among other things. Readers not fully comfortable with these concepts are strongly encouraged to consult other resources, such as the appendix to Griffiths [1, p. 435]. A general understanding of basic calculus (including derivatives, partial derivatives, integrals, and differential equations) is needed for a good mathematical understanding of Sections 1.3 and 3.3, but is otherwise unnecessary.

This part of the paper does not come even close to being a comprehensive survey of quantum theory and its subfields. Readers that are interested in learning more about the topic are invited to consider further reading such as (ordered roughly by approachability) Hughes [2], Nielsen and Chuang [3], Griffiths [1], and Sakurai and Napolitano [4]. Hughes provides a mathematical account of quantum foundations and philosophical discussions. Nielsen and Chuang provide a comprehensive survey of quantum computation and information theory. This is a good framework for beginners since the systems of study are considerably simpler than those in pure quantum physics. Griffiths's book is the standard undergraduate text on quantum mechanics, whereas advanced or graduate level courses often choose the text

by Sakurai and Napolitano. Chemists or readers interested in an old-school physics-heavy presentation of quantum mechanics should consult Pauling and Wilson [5].

Readers interested in tackling quantum field theory should begin by becoming comfortable with the above resources. Then, they should become familiar with the path-integral approach of quantum mechanics as presented by Feynman [6]. The next resource I would suggest is the elementary particles book by Griffiths [7]. Finally, such a reader would be equipped to move on to proper quantum field theory texts. Zee [8] is a good place to start.

On a more stylistic note, throughout this paper I will address the reader directly, parenthetically, and in footnotes. My hope is that this conversational tone will help the reader to grasp the topics at hand. I encourage the reader to explore avenues I leave open, ponder ideas I introduce, fill in proofs and exercises I have omitted, and consult the references I cite. I want this paper to serve as a jumping-off point that can be used in a very practical way.

Chapter 1

The Postulates of Quantum Theory

Quantum theory creates a new paradigm for thinking about and formalizing systems. Depending on the details of how the theory is applied, from it one can develop quantum mechanics, quantum field theory, quantum information science, quantum probability theory, and many other fields. The concepts that these fields share are collectively known as quantum theory. In this chapter, I develop the framework of quantum theory by introducing four working postulates: the state space postulate, the measurement postulate, the evolution postulate, and the composite system postulate.

1.1 The State Space Postulate

Quantum theory concerns itself with systems and states of these systems. The state space postulate asserts that the state of any system is represented by a vector in a complex Hilbert Space \mathcal{H} . That is to say, a complex vector space equipped with an inner product. Mathematicians denote the inner product by $\langle \cdot, \cdot \rangle$ which is understood to be linear in the left-hand term and antilinear in the right-hand term. For reasons that will soon become apparent, physicists use the notation $\langle \cdot | \cdot \rangle$ which is understood to be linear in the *right-hand* term and

antilinear in the left-hand term.

Adopting a wonderful notation introduced by Dirac, we denote each state by a “ket” $|\psi\rangle$ where ψ is some label. These kets can be thought of as column vectors. The dual vector to a ket $|\psi\rangle$ is denoted by a “bra” $\langle\psi|$. These bras¹ can be thought of as row vectors, found by taking the complex conjugate and transpose of the associated column vector². Thus, the inner product of two vectors labeled by ψ and ϕ is given by $\langle\phi|\psi\rangle = \langle\phi|\psi\rangle$, which served as the inspiration for Dirac’s notation.

When working in finite dimensions, kets are quite literally column vectors and bras are the corresponding row vectors. In this setting, we can define a basis $|0\rangle, |1\rangle, \dots$ with the property $\langle i|j\rangle = \delta_{ij}$. Such a basis is known as an orthonormal basis, and these orthonormal bases are the foundation of the column vector, row vector, matrix notation commonly seen in linear algebra.³ In this context, we can always expand a ket by

$$|\psi\rangle = \sum_k \langle k|\psi\rangle |k\rangle = \sum_k \psi_k |k\rangle \quad (1.1)$$

where $\langle k|\psi\rangle = \psi_k$ are the column vector coefficients. Likewise, row vectors can be written out as

$$\langle\psi| = \sum_k \langle\psi|k\rangle \langle k| = \sum_k \psi_k^* \langle k| \quad (1.2)$$

where $*$ denotes complex conjugation. As suggested by the notation, the inner product found

¹For the curious mind, yes, this terminology finds itself at the end of many a physicist’s joke.

²We denote the operations of complex conjugation together with transposition by \dagger . So, we have $\langle\psi| = |\psi\rangle^\dagger$ and vice versa.

³Given some orthonormal basis $|i\rangle$, a linear operator L can be represented by a matrix whose elements in the $|i\rangle$ basis are given by $L_{ij} = \langle i|L|j\rangle$. The columns of the matrix represent what vectors the basis vectors get sent to when operated on. The usual rules for matrix multiplication and addition follow from this. Likewise the notation for row and column vectors comes from this approach. If you were never taught about matrices in this way, I suggest you read Section A.3 in the appendix of Griffiths [1, p. 441]. Understanding this is crucial to developing a mathematical intuition for quantum theory.

by simply matrix multiplying the row vector and the column vector:

$$\langle \phi | \psi \rangle = \sum_k \phi_k^* \psi_k. \quad (1.3)$$

In infinite dimensions, we often use bases that are not countable. In this context we appeal to continuous versions of the above equations:

$$|\psi\rangle = \int_{-\infty}^{\infty} \langle x | \psi \rangle |x\rangle dx = \int_{-\infty}^{\infty} \psi(x) |x\rangle dx \quad (1.4)$$

$$\langle \psi | = \int_{-\infty}^{\infty} \langle \psi | x \rangle \langle x | dx = \int_{-\infty}^{\infty} \psi^*(x) \langle x | dx \quad (1.5)$$

$$\langle \phi | \psi \rangle = \int_{-\infty}^{\infty} \phi^*(x) \psi(x) dx \quad (1.6)$$

where $\psi(x)$ is often called the “wavefunction”. Even when we work in this picture it is still useful, perhaps even crucial, that one thinks of kets like column vectors, bras like row vectors, $\psi(x)$ like a coefficient, and so on. In quantum mechanics, it is always possible to define a countable (though often not finite) orthonormal basis. Thus, we could always switch to a basis where all this is clearer, but one rarely does; therefore, make sure you understand the correspondences between the countable and uncountable cases.

For reasons that will become clear in the next section, we insist that all physical states be represented by *normalized* vectors in the Hilbert space for our system. This forces the relation

$$\langle \psi | \psi \rangle = 1 \quad (1.7)$$

which in a continuous basis looks like

$$\int_{-\infty}^{\infty} \psi^*(x) \psi(x) dx = \int_{-\infty}^{\infty} |\psi(x)|^2 dx = 1. \quad (1.8)$$

Elements in a continuous basis often can't be normalized; thus they can't represent true

physical states. However, they are often useful to work with. The best we can do for a continuous basis is to enforce the so-called Dirac normalization $\langle x|x' \rangle = \delta(x - x')$, which can be thought of as the continuous analog of the orthonormality condition $\langle i|j \rangle = \delta_{ij}$. In the former, case we use the Dirac delta function; in the latter, we use the Kronecker delta symbol.

Note that Equation 1.7 allows $|\psi\rangle$ to be multiplied by an arbitrary complex number of modulus 1, i.e. some $e^{i\theta}$, without breaking the constraint. We call such a number a “global phase”. Fortunately, global phases have no physical consequences⁴, and we don’t need to further constrain our states. In fact, one might go as far as to say that the collection of all of those kets that vary by nothing more than a global phase collectively represent the state of the system. However, physicists are by definition not this pedantic. We call those who are “mathematicians”.⁵

1.2 The Measurement Postulate

The measurement postulate asserts that all observable quantities (such as position and momentum) can be represented by Hermitian⁶ operators that act on the Hilbert space of the system. By the spectral theorem, we can decompose these operators into a sum over their eigenvalues and corresponding eigenspace projections:

$$O = \sum_i \lambda_i P_{\lambda_i} \tag{1.9}$$

⁴Despite this fact, *relative* phases *do* matter. A state $|\kappa_1\rangle = \frac{1}{\sqrt{2}}(|\phi\rangle + e^{i\theta_1}|\psi\rangle)$ is most definitely different from a state $|\kappa_2\rangle = \frac{1}{\sqrt{2}}(|\phi\rangle + e^{i\theta_2}|\psi\rangle)$ when $\theta_1 \neq \theta_2$.

⁵Clearly you are one of them if the wording of that sentence is bothering you in light of the sentence preceding it.

⁶Hermitian operators are linear operators O for which $O = O^\dagger$ holds.

where O is the operator⁷ in question, the λ_i 's are its eigenvalues, and P_{λ_i} is the projection operator that projects onto the subspace spanned by all eigenvectors that have eigenvalue λ_i . For Hermitian operators, we are guaranteed that these eigenvalues are real numbers and that the projections are pairwise orthogonal. Moreover, we can always find an orthonormal basis of eigenvectors so that

$$O = \sum_i \lambda_i |\lambda_i\rangle\langle\lambda_i|. \quad (1.10)$$

As is commonly done in physics, I have denoted the projection onto the vector $|\lambda_i\rangle$ by $|\lambda_i\rangle\langle\lambda_i|$. This follows from the fact that

$$|\psi\rangle\langle\psi| |\phi\rangle = |\psi\rangle\langle\psi|\phi\rangle = \langle\psi|\phi\rangle |\psi\rangle. \quad (1.11)$$

In fact, depending on how you look at it, this is more than just a notational convention. In this expansion⁸ of O , it is possible that there might be two or more eigenvalues that are equal so that O is “degenerate”. In which case, the sum over the corresponding $|\lambda_i\rangle\langle\lambda_i|$ projections gives the eigenspace projection P_{λ_i} .

The measurement postulate asserts that the eigenvalues represent the possible values for the observable and that the normalized eigenvectors represent the states of the system that have a well defined value for that particular observable. If one measures a system in an eigenstate of the observable in question, the system will remain unchanged and the corresponding eigenvalue will be the measurement result.

⁷I could use a hat or other notation to distinguish operators, but authors in physics are notorious for lacking the foresight to make such distinctions, so it is worth conforming to this standard if for no other reason than to give the reader practice. I will try to restrict capital letters to operators to prevent confusion.

⁸It is worth taking a quick second here to note that the identity operator I , for which every state is an eigenvector with eigenvalue 1, can be decomposed into any basis (discrete or continuous) via:

$$I = \sum_k |k\rangle\langle k| = \int_{-\infty}^{\infty} |x\rangle\langle x| dx \quad (1.12)$$

or any analogous formula.

What about systems that are *not* in an eigenstate of the observable being measured?⁹

Here the measurement postulate¹⁰ asserts that for a system in¹¹ a state $|\psi\rangle$, upon measuring O , one will find the measurement result to be λ_i with probability $\langle\psi|P_{\lambda_i}|\psi\rangle$ and the state of the system will subsequently “collapse” onto the state $\frac{P_{\lambda_i}|\psi\rangle}{\sqrt{\langle\psi|P_{\lambda_i}|\psi\rangle}}$. The denominator simply ensures normalization. If the observable is non-degenerate, then this procedure reduces to returning λ_i with probability $\langle\psi|\lambda_i\rangle\langle\lambda_i|\psi\rangle = |\langle\psi|\lambda_i\rangle|^2$ and collapsing onto the state $|\lambda_i\rangle$. This rule for calculating probabilities is known as Born’s rule.

Such general states, which can always be written as a linear combination (or “superposition”) of eigenstates since they form a basis, do *not* have well-defined values of the observable in question. It is like asking what color a song is; the question does not make sense in context. Often popular science accounts will treat such superpositions as simultaneously having all the values of the eigenstates it includes, but this is a very misleading and dangerous metaphor. It is best to think of such states as simply being different states of the system for which the observable isn’t well-defined and that happen to be given by combinations of states for which the observable is defined.

This is a very important point so it’s worth emphasizing once more. Popular science accounts will talk about particles that are in a superposition of two well-defined position states as if they are in *both* places at once.¹² This is nonsense. Asking “Where is the particle?” when the particle is in a superposition of position states is like someone playing

⁹You can apply the procedure I outline in this paragraph to eigenstates as well. What you’ll find is that one probability is 1 and the rest are 0, making this procedure reduce to the simpler one discussed in the previous paragraph.

¹⁰It should be noted here that I am only discussing so-called von Neumann measurements. There are other formalisms for measurements in quantum mechanics, such as POVM’s, that I do not discuss. As I discuss in Section 1.3, the exact details of measurement in quantum theory is an open question.

¹¹It is technically wrong to speak of a system in a state $|\psi\rangle$. It is in a state *represented* by $|\psi\rangle$. This paper, like the rest of the physics community, will rarely draw attention subtleties like this.

¹²Some go even further and say that physicists have actually seen particles be in two places at once. There should be a prison sentence for such remarks. It’s not that they are making quantum physics seem more magical than it really is. It’s that they are making it seem *less* magical than it really is. Not to mention they are blatantly lying about what we have and haven’t seen in the lab.

a musical chord and asking “What note is this?” The question cannot be applied because the chord isn’t a note. Say the chord is made by playing two notes, C and E, at the same time. The chord is not “the note C”; nor is it “the note E”. It is most certainly not *both* the note C and the note E. It is the result of the note C and the note E being *played at the same time*. Now, return to the particle. We can think of “the particle is here” as analogous to “the note C” and “the particle is there” as analogous to “the note E”. A particle in a superposition of being here and there is analogous to the chord we considered.¹³ Such a particle is neither here nor there. It is most certainly not *both* here and there. Instead, the *state* of the particle is the result of “the particle is here” and “the particle is there” *being played at the same time*, so to speak. If one could develop an intuition for what it meant for “the particle is here” and “the particle is there” to be “played at the same time”, one would truly *understand* quantum theory. This is the essential mystery of quantum theory.¹⁴

Returning to the math, the point of forcing kets to be normalized becomes clear here. Normalization ensures that the probabilities given by Born’s rule sum up to 1 for any observable. Furthermore, this allows us to interpret the coefficients of $|\psi\rangle$ in any basis to be “amplitudes” whose magnitudes squared give the probabilities associated with collapsing onto the corresponding basis states (and returning the corresponding eigenvalues) after some measurement. For example, in the case of the one-particle one-dimensional position operator, defined by:

$$X = \int_{-\infty}^{\infty} x |x\rangle\langle x| dx \quad (1.13)$$

the magnitude squared of the wavefunction $|\psi(x)|^2$ represents the probability of finding the particle at position x . Of course, as we remarked earlier, these $|x\rangle$ states can’t be normalized,

¹³This analogy is much deeper than one might expect. One can represent a position eigenstate by a plane wave in the momentum basis, where different positions are mapped to different frequencies. Then, the superposition of two such states is given by literally *adding* these two waves of different frequencies. This is exactly the same way that a chord forms: by adding the waves formed by two notes in open air.

¹⁴Well, at least one would understand the mystery of superpositions. The “measurement problem” (discussed briefly in the next section) is another story.

so one can't really perform a perfect X measurement. However, discussing things in these loose physical terms is useful to gain intuition for the mathematics.

A useful tool for analyzing observables is the commutator, defined for two observables O_1 and O_2 by

$$[O_1, O_2] = O_1 O_2 - O_2 O_1 \quad (1.14)$$

which is equal to 0 in the case that the two operators commute. When in fact two observables commute, it is a mathematical fact that they share a complete¹⁵ set of common eigenvectors. It follows from this and the measurement postulate that the order in which one performs two measurements has the same statistical results when the two corresponding observables commute. Thus, we say that commuting observables are simultaneously measurable. When both are non-degenerate, the two only differ in their eigenvalues and are thus essentially the same measurements with relabeled outcomes. It turns out that the converse is also true: any two operators that share a complete set of common eigenvectors (and thus whose measurement statistics are independent of order) necessarily commute. Therefore, being simultaneously measurable is equivalent to commuting. Operators that don't commute *cannot* be simultaneously measured because the order in which one performs the measurements changes the statistics; the statistics change because of details concerning how the state collapses twice onto two different bases. Non-commuting operators have further consequences that I will discuss in Chapter 2.

To summarize, the measurement postulate says that each observable corresponds to a Hermitian operator. This operator should be thought of as a little suitcase that contains all of the information needed to do measurements of that observable. Just as it is usually pointless to smack things with a suitcase, it is usually meaningless to apply this operator directly to a state. The spectral theorem (see Equation 1.9) shows us that the suitcase can be unpacked into eigenvalues and eigenvectors. The eigenvalues are measurement outcomes

¹⁵By complete we mean that the set forms a basis for the Hilbert space.

and the eigenvectors are the states the system will collapse onto with certain probabilities. However, unlike a real suitcase, one can find some use to these operators without unpacking them. For example, it is easy to show¹⁶ that the statistical average of the measurement outcome for an observable O measured on a system in state $|\psi\rangle$ is given by $\langle\psi|O|\psi\rangle$. We often denote this by $\langle O \rangle_\psi$ or more often by $\langle O \rangle$ where the state dependence is left implicit.

1.3 The Evolution Postulate

Determining how the state of a system changes with time is perhaps the most important role of physics in general. This is precisely what the evolution postulate is concerned with. Until now, we have not shown states to have any time dependence, but presumably the state of a system changes over time and it would be more accurate to reflect that by using a notation like $|\psi(t)\rangle$. From time to time I will continue to leave time dependence implicit.

The evolution postulate takes many forms, often depending on what field you are working in, whether it be quantum mechanics, quantum field theory, or some other field. All of these postulates are equivalent in principle, provided one is careful enough. Perhaps the most well-known form used in quantum mechanics is the famous time-dependent Schrödinger equation

$$i\hbar \frac{\partial}{\partial t} |\psi\rangle = H |\psi\rangle \quad (1.15)$$

where H is the Hamiltonian (total energy) operator and $\hbar = \frac{h}{2\pi} = 1.054\,572\,726(47) \times 10^{-34}$ J s is Planck's reduced constant. Solving this equation, one finds that the evolution of the state is given by

$$|\psi(t)\rangle = U(t, t_0) |\psi(t_0)\rangle = e^{-i\frac{H}{\hbar}(t-t_0)} |\psi(t_0)\rangle \quad (1.16)$$

where U is the evolution operator or “propagator”.¹⁷ The obvious question you should be

¹⁶Hint: Use the spectral theorem!

¹⁷The solution I show here is only valid when the Hamiltonian does not explicitly depend on time. This is

asking yourself is: “How do I take e to the power of an operator?” The question is answered by a branch of functional analysis known as functional calculus. You know how to take powers of operators, right? And you know how to expand most functions, such as e^x , into a power series, right? The short answer, then, is that we can input operators into arbitrary functions by inserting them into the Taylor series expansion and summing over the various powers of the operator.¹⁸ Still, this seems a bit tedious. Fortunately, due to the spectral theorem it is possible to show that for an operator O expanded by Equation 1.9 and most single variable functions f , this procedure results in

$$f(O) = \sum_i f(\lambda_i) P_{\lambda_i} \quad (1.18)$$

i.e. we leave the eigenvectors alone and just evaluate the function on the eigenvalues.

In one particle, one-dimensional quantum mechanics, it is common to use the Hamiltonian $\frac{P^2}{2m} + V(X)$ where X and P are the position and momentum operators, m is the mass of the particle, and V is some potential energy function. This, in addition to the relation

$$\langle x | P | \psi \rangle = -i\hbar \frac{\partial}{\partial x} \langle x | \psi \rangle \quad (1.19)$$

which shows how the P operator looks in the $|x\rangle$ basis (and whose proof is beyond the scope of this paper) leads to the familiar form of the Schrödinger equation

$$i\hbar \frac{\partial}{\partial t} \psi(x, t) = \frac{-\hbar^2}{2m} \frac{\partial^2}{\partial x^2} \psi(x, t) + V(x) \psi(x, t) \quad (1.20)$$

quite often the case, but don’t get fooled into thinking that the final equality in Equation 1.16 is generally equivalent to—or more fundamental than—the Schrödinger equation. When the Hamiltonian depends on time, $U(t, t_0)$ does not take the simple form $e^{-i\frac{H}{\hbar}(t-t_0)}$. In general, $U(t, t_0)$ satisfies the equation

$$i\hbar \frac{\partial}{\partial t} U(t, t_0) = HU(t, t_0). \quad (1.17)$$

¹⁸Longer answers deal with the case where we can’t appeal to such a series.

which shows how the $|x\rangle$ basis coefficients of $|\psi\rangle$ evolve.

Also familiar to most people is the time-independent Schrödinger equation

$$H |\psi\rangle = E |\psi\rangle \quad (1.21)$$

which is perhaps less misleadingly written as

$$H |\phi_n\rangle = E_n |\phi_n\rangle. \quad (1.22)$$

This is simply the eigenvector equation for the Hamiltonian, where the E_n are real numbers corresponding to eigenvalues which represent measurable values of energy. It is typically used for finding what the energy eigenstates look like in some already chosen basis (such as the position basis) for which the Hamiltonian's form has been discerned, or for finding out what the possible energy values are. One of the most fundamental results of quantum mechanics is that it isn't usually possible for a system to have any arbitrary energy. Typically, the energy can only take very specific, discrete values because the Hamiltonian itself usually has very specific, discrete eigenvalues.

Applying the propagator to an energy eigenstate we find

$$|\phi_n(t)\rangle = U(t, t_0) |\phi_n(t_0)\rangle = e^{-i\frac{E_n}{\hbar}(t-t_0)} |\phi_n(t_0)\rangle \quad (1.23)$$

where $\frac{E_n}{\hbar}$ is just a number now, which we often call ω_n . As you can see, energy eigenstates only pick up a global phase over time; thus, we call them stationary states.¹⁹ If we write out a general state in terms of the energy eigenbasis, we find that the coefficients evolve in time by picking up a phase related to the energy of each basis state.²⁰ From all this (and our choice of letter), perhaps you have noticed by now that U is in fact a unitary operator,

¹⁹Remember, global phases have no physical consequences.

²⁰Remember, relative phases *do* have physical consequences.

i.e. $U^\dagger = U^{-1}$. In fact, e^{-iO} for any Hermitian operator O is a unitary operator. The fact that the propagator is unitary is crucial: it guarantees that the state normalization will be preserved over time, since

$$\langle \psi(t) | \psi(t) \rangle = \langle \psi(t_0) | U^\dagger(t, t_0) U(t, t_0) | \psi(t_0) \rangle = \langle \psi(t_0) | I | \psi(t_0) \rangle = \langle \psi(t_0) | \psi(t_0) \rangle = 1. \quad (1.24)$$

Essentially, Schrödinger's equation says that the states evolve in time by rotating around in the Hilbert space of the system. Often, we may pass the state of the system through some “gate” that applies a known Hamiltonian for a fixed amount of time. In these cases, it is easiest to specify the unitary operator U that transforms the state before the gate into the state after the gate.

This all amounts to what is known as the Schrödinger picture of evolution. Operators remain fixed (unless explicitly time-dependent, which is rare), and states rotate around in the Hilbert space according to the propagator. However, we could equivalently hold all of the states fixed and rotate the operators instead. This corresponds to the Heisenberg picture of evolution. Given a Schrödinger picture operator A_S , we define its Heisenberg picture equivalent by

$$A_H(t) = U^\dagger(t, 0) A_S U(t, 0). \quad (1.25)$$

This definition ensures that the measurement postulate takes the same form as before for a given time t , and subsequently that expectation values are still calculated by surrounding the operator by a bra and ket. From this definition, it is easy to derive²¹ the Heisenberg equation of motion

$$\frac{d}{dt} A_H(t) = \frac{1}{i\hbar} [A_H(t), H] + \left(\frac{\partial A_S}{\partial t} \right)_H. \quad (1.26)$$

²¹Hint: Use the product rule, then use Equation 1.17 and its complex conjugate (don't forget to switch the order of H and U when complex conjugating). Don't worry about the difference between total and partial time derivatives. The partial derivative is used in Schrödinger's equation to emphasize that it is only a time derivative. The distinction becomes important when considering the evolution of the wavefunction (the $|x\rangle$ basis coefficients) $\psi(x, t)$ as in Equation 1.20.

Don't confuse the subscript H for the Heisenberg picture with the Hamiltonian H . The last term is almost always zero, as very few operators are explicitly time-dependent in the Schrödinger picture.²² Students of classical mechanics will realize that this is the same equation as the general form of Hamilton's equation if we replace the commutator over $i\hbar$ with the Poisson bracket. It turns out that indeed this is almost²³ always the right correspondence between classical Poisson brackets and quantum mechanical commutators. Thus, in this context we can conclude that the Heisenberg equation of motion is essentially the same as the Hamilton's equations for the observable in question. For this reason, many see the Heisenberg picture as the more fundamental one. It also turns out that this picture can be applied in a more straightforward way to deeper studies (such as quantum field theory) than the Schrödinger picture. However, it lacks the intuitiveness of the Schrödinger picture in which the state of the system evolves in time rather than the observables. Fortunately, as is often the case in physics, we get the best of both worlds because they are equivalent. We can use whichever suits our particular need at the time.

It turns out that there is yet another formulation of the evolution postulate due to Feynman and Dirac. Known as the path-integral formulation, it dictates that the state evolves according to some propagator $U(t, t_0)$ as before. Now, however, the question of

²²At this point you've seen enough quantum theory to appreciate a *real* quantum joke, not some lame one about Schrödinger's cat walking into a bar. One day a caveman Zog, troubled about his health, goes to visit his doctor:

ZOG: Doc, I'm getting worried. Everyone except for me seems to be changing in weird ways.

DOCTOR: Hmm. How are they changing?

ZOG: Well, their backs are getting straighter and they are losing body hair.

DOCTOR: Maybe it's something to do with where you work.

ZOG: I don't think so. My friends at work are changing too.

DOCTOR: How do you get to work?

ZOG: I carpool with my friend Thug. Come to think of it, he's not changing either.

DOCTOR: Hmm... and I assume Thug lives in Kroktown with you?

ZOG: Actually no, he lives over in Hamilton.

DOCTOR: Ah, well there's your problem! You'll never evolve if you commute with the Hamiltonian.

²³This correspondence does not *always* hold in general. However, there is a correspondence between the quantum commutator and a deformation of the Poisson bracket known as the Moyal bracket. The formalization of this correspondence is known as the Groenewold-Van Hove theorem. See Groenewold [9].

interest is how to find what this propagator looks like in a particular basis (usually the position basis). The result that Feynman came up with is that

$$\langle x' | U(t, t_0) | x \rangle = \int e^{i \frac{S[x(t)]}{\hbar}} Dx(t) = \int e^{i \frac{\hbar}{\hbar} \int_{t_0}^t \mathcal{L} dt} Dx(t). \quad (1.27)$$

That is to say, the amplitude for a particle to travel from x to x' in the time interval t_0 to t is given by integrating over every path the particle could take to get there and assigning each path an amplitude where the classical action \mathcal{S} of the path is to be interpreted as a phase! The amount of beauty in this formulation cannot be stated in mere words. This is perhaps *the* most fundamental formulation of the evolution postulate. It is by far the easiest to carry into quantum field theory, and it is arguably the best formulation for teasing out how classical physics emerges from quantum mechanics. Unfortunately, to expand upon its beauty would be well beyond the scope of this paper, and I urge you to study this topic further elsewhere.²⁴

Even more unfortunately, this formulation often proves to be the most difficult to work in for quantum mechanics, especially if one is concerned with mathematical formalism. As you can imagine, integrating over every possible path is no easy matter. Quite often it is impossible to carry out the integral. Fortunately, this formulation can be proven to be equivalent to the others, so, if we wish, we can merely remark on its beauty and use some other picture for calculating results. However, in quantum field theory, it turns out quite often that this picture is far easier to get physical results from—which just goes to show what a monster of a field QFT is.

One thing that I have failed to remark on so far is the apparent contradiction between the various evolution postulates and the measurement postulate. The evolution postulate is

²⁴For a wonderful derivation of this formulation, see the first chapters of Zee [8]. For an entire (and excellent) book on this formulation and how to apply it to simple quantum mechanics (as opposed to quantum field theories), see Feynman [6]. For the fascinating history of how Feynman came up with it, and what part Dirac played, see Feynman's Nobel Prize lecture [10].

strictly continuous and deterministic. In principle, it ought to apply to any closed quantum system. Thus, if one considers the total system of some particle together with a measuring apparatus and the experimenter operating it, then one would expect that the total quantum state would evolve continuously and deterministically. However, the measurement postulate dictates random discontinuous collapses in the state (well, at least the substate corresponding to the particle)! How can we reconcile these two? If I had a universally satisfying answer to this question, I'd probably be writing a Nobel prize lecture right now instead of an undergraduate thesis.²⁵ This “measurement problem” is arguably the largest open question in quantum mechanics and its interpretations. I encourage you to dive down the rabbit hole, read up on the tomes of history surrounding it, and let it bother you as much as possible. However, this poses another question which I *can* tell you a little about. That is, how do you combine subsystems into a total system with a total state?

1.4 The Composite System Postulate

The composite system postulate asserts that the Hilbert space for a larger system containing two smaller systems, whose Hilbert spaces are \mathcal{H}_1 and \mathcal{H}_2 , is given by the tensor product space $\mathcal{H} = \mathcal{H}_1 \otimes \mathcal{H}_2$. If you have never studied tensor products, try not to worry too much about the formal details. It is essentially just a way of keeping track of two vector spaces at the same time. The tensor product of a m dimensional space and a n dimensional space is a $m \times n$ dimensional space. If we have a basis $|u_i\rangle$ for \mathcal{H}_1 and a basis $|v_i\rangle$ for \mathcal{H}_2 , then the set of tensor products of these states $|u_i\rangle \otimes |v_j\rangle$ is a basis for \mathcal{H} . Whether it be the tensor product of spaces, vectors, or operators, the most important thing to keep in mind about the tensor product is that it is linear in both terms.

When the total system is in a “product state” $|\psi\rangle \otimes |\phi\rangle$, it is understood that the first

²⁵At the very least this paper would have a different title.

subsystem is in the state $|\psi\rangle$ and the second subsystem is in the state $|\phi\rangle$. A general bipartite state can be written as a linear combination of these product states. States that cannot be written as some single product state are called “entangled”. We will discuss such states in Chapter 4.

The observable corresponding to measuring observable O_1 on the first subsystem and O_2 on the second subsystem is given by the operator $O = O_1 \otimes O_2$. The tensor product of two operators is defined by the equation

$$(O_1 \otimes O_2)(|\psi\rangle \otimes |\phi\rangle) = O_1 |\psi\rangle \otimes O_2 |\phi\rangle. \quad (1.28)$$

The observable corresponding to measuring O_1 on the first subsystem and *not* measuring the second subsystem is given by the operator $O = O_1 \otimes I$. Likewise, we can construct local observables for the second subsystem. A general observable for the total system can be written as linear combinations of these product operators. Just as for states, observables that cannot be written as simple products are called entangled or entangling observables. These are observables that have at least one entangled eigenstate.

The inner product, like everything else for composite systems, is defined by how it works for product states:

$$(\langle\phi_1| \otimes \langle\phi_2|)(|\psi_1\rangle \otimes |\psi_2\rangle) = \langle\phi_1|\psi_1\rangle \langle\phi_2|\psi_2\rangle. \quad (1.29)$$

If you express everything in the first system (states and operators) in matrix form for some basis, and everything in the second system in matrix form for some other basis, then the tensor products of states and operators in matrix form for the tensor product basis is given by interpreting \otimes as a Kronecker product²⁶. In this case, Equation 1.29 falls out naturally.

²⁶I realize you probably don’t know how to take the Kronecker product of matrices, so this is far from being the “help” I intend for it to be in understanding how the tensor product works. I apologize for not spelling it out here as it takes big spaces and isn’t crucial to my work. You can find it in any popularly accessible resource, such as Wikipedia. I encourage you to derive it from the properties of tensor products that I have shown and the general way of representing vectors and operators as matrices in a given basis.

One can represent evolutions and gates as tensor products of local unitaries, and one can build global unitaries out of these that cannot be represented by a simple single tensor product of unitaries. Such gates are called entangling gates. The Hamiltonian of a joint system is the tensor product of the Hamiltonians of the two local systems, provided they don't interact. However, this is rarely the case²⁷. We usually have to form a Hamiltonian that includes interactions between the two systems and subsequently can entangle them as they evolve in time. There is not much else to say about the composite state postulate. However, I should make a quick note about notation. Product states like $|\psi\rangle\otimes|\phi\rangle$ are sometimes notated as $|\psi\rangle|\phi\rangle$ or even $|\psi,\phi\rangle$. Some authors also take care to subscript each bra or ket with a label corresponding to which system it applies to, like $|\psi\rangle_A\otimes|\phi\rangle_B$. I won't swear by any one notation, and you'll find a mixture of all of them in the literature. What I *will* swear by is that the left part of the product always corresponds to the first subsystem, and the right part of the product always corresponds to the second subsystem. Some authors reverse this for the case of bras so that $(|\psi\rangle|\phi\rangle)^\dagger|\psi\rangle|\phi\rangle = \langle\phi|\langle\psi||\psi\rangle|\phi\rangle = \langle\phi|\langle\psi|\psi\rangle|\phi\rangle = \langle\psi|\psi\rangle\langle\phi|\phi\rangle$. This seems nice at first, until you realize that everything else about the tensor product is the other way around and begin to confuse the two notations. In fact, the whole spirit of the tensor product is that it concatenates two systems, so we ought to be consistent. Fortunately, I am far from alone in my convention. Just be on the lookout for this divergence when reading other sources.

The Kronecker product of vectors is just that for n by 1 (or 1 by n) matrices.

²⁷If it were always the case, why would we care about the joint system in the first place!

Chapter 2

The Uncertainty Principle

In addition to the idea that linear combinations (superpositions) can create particles that are “in two places at once”, the uncertainty principle joins the ranks of the worst misrepresented quantum concepts in popular science. We have all heard some high school chemistry teacher tell us, “You can’t know exactly where an electron is and how fast it’s moving at the same time.” The question you have to ask yourself is why you’d take advice about physics from a high school chemistry teacher in the first place... especially advice about *quantum* physics. The uncertainty principle has a very precise mathematical meaning, and we will explore it in this chapter¹.

2.1 The Old Uncertainty Principle

As the story goes, Heisenberg figured out that non-commuting observables could not be measured simultaneously. Furthermore, the more you knew about one observable, the less you knew about the other, and vice versa. He knew that non-commuting observables obeyed

¹This chapter isn’t crucial to my work, but you can’t write about the foundations of quantum mechanics without mentioning the uncertainty principle. To do so would be a crime. However, to write about the uncertainty principle in a *wrong* way—or even worse, a *misleading* way—would be a greater crime. Hence this chapter. You don’t have to read it if you don’t want to.

a sort of uncertainty relation, but he never made this relation entirely precise. What he did come up with, was that uncertainties in position and momentum are related by something like

$$\Delta x \Delta p \gtrsim \hbar. \quad (2.1)$$

His reasoning followed from applying Fourier optics principles to de Broglie's concept of matter waves, but without giving precise definitions of what these Δ 's meant. When he explained the concept, he often used a thought experiment in which a particle's position is measured by illuminating the particle with light (or other radiation) and then observing the reflected light using a microscope. In order to measure the particle's position precisely, one needs to use high frequency radiation. However, high frequency radiation greatly disturbs the particle's momentum.

Heisenberg [11, 12] gave rough mathematical arguments to justify the relation, but never gave a completely formal proof. Kennard [13] and Weyl [14] are credited with proving the more modern relation

$$\sigma_x \sigma_p \geq \frac{\hbar}{2} \quad (2.2)$$

(where the σ 's represent statistical standard deviations) shortly after Heisenberg's original paper. In any event, Heisenberg's primitive relation eventually inspired Robertson [15] to come up with the much more precise and general relation we now know as the modern uncertainty principle.

2.2 The Modern Uncertainty Principle

The modern uncertainty principle due to Robertson is given by

$$\Delta_\psi A \Delta_\psi B \geq \left| \frac{1}{2i} \langle [A, B] \rangle_\psi \right| \quad (2.3)$$

where $\Delta_\psi A = \sqrt{\langle\psi|(A - \langle A\rangle_\psi)^2|\psi\rangle}$ is the statistical standard deviation in the measurement outcomes of A when the system is in state $|\psi\rangle$, and both A and B are understood to be Hermitian.² As you can see, this modern relation is just a statement about the measurement statistics for non-commuting (or non-anticommuting) operators. In some sense, it puts bounds on how well-defined two observables can be for a given state $|\psi\rangle$.³

The key to all this is that both sides of the relation are dependent on the state: each state specifies a different uncertainty relation. The average of the commutator might even be zero for some state. In that case, the (Robertson) uncertainty relation for such a state is trivial, since the left side is obviously greater than zero. Furthermore, it's even possible that this average is zero for a state *because* that state is a simultaneous eigenstate of both operators. It isn't possible to construct a complete set of common eigenvectors for non-commuting operators, but it is easy to find two observables in as few as 3 dimensions that don't commute yet share at least a single eigenvector. Beyond this, the right hand side necessarily goes to zero for a state that is an eigenstate of *either* operator. This condition forces the corresponding standard deviation to be zero and subsequently causes the entire left-hand side product to be zero. It is also imaginable that we could construct some state and observables such that the state is not an eigenstate of either operators, yet the average of the commutator still goes to zero. I won't speculate on whether or not this is mathematically

²A simple proof of this relation can be found in Section 3.5 of Griffiths [1, p. 110]. However, I urge you to try your own hand at its proof. Begin by applying the Cauchy–Bunyakovsky–Schwarz inequality $\langle x, x \rangle \langle y, y \rangle \geq |\langle x, y \rangle|^2$ to vectors like $x = (A - \langle A \rangle_\psi)|\psi\rangle$ and $y = (B - \langle B \rangle_\psi)|\psi\rangle$. This should give you an uncertainty relation for variances. Then, expand the right side in terms of its real and imaginary parts. Throw away the real part (justify why you can, of course!), play around a bit, then take the square root of both sides, and you'll end up with Equation 2.3. If you're up for it, try keeping both the real and imaginary parts and you can come up with the more general, stronger relation that Schrödinger later proved:

$$\Delta_\psi A \Delta_\psi B \geq \sqrt{\left(\frac{1}{2} \langle \{A, B\} \rangle_\psi - \langle A \rangle_\psi \langle B \rangle_\psi\right)^2 + \left(\frac{1}{2i} \langle [A, B] \rangle_\psi\right)^2} \quad (2.4)$$

where $\{A, B\} = AB + BA$ is the anticommutator.

³As we noted before, observables are only well-defined for eigenstates, but this particular relation refers to a statistical way of looking at how close the state is to being well-defined for each observable.

possible, but this whole topic should serve to emphasize how crucial (if not unfortunate) it is that the right-hand side is state dependent. The punchline is that it prevents the uncertainty principle from being some grand statement limiting the measurement statistics of observables objectively, on their own groundings, independent of the details of the state of the system.

But wait! What if $[A, B]$ is some constant multiple of the identity I ? Then the expectation value would be the same for *any* state. It would just be the constant, which doesn't depend on the state. It turns out this type of commutator only occurs for special infinite-dimensional operators. As a matter of fact, it turns out that $[X, P] = i\hbar I$. So we have

$$\Delta X \Delta P \geq \frac{\hbar}{2} \quad (2.5)$$

where we have suppressed the dependence of the left side on the state because it is obvious and the right hand side is state independent.

What can we say now about our old chemistry teachers' claims? Well, according to quantum theory, there is no state whose position and momentum are both well-defined⁴. If you believe quantum theory is a complete description of the world, then the terms position and velocity never both make sense when applied to an electron at the same time. How much statistical sense they can both make at the same time is bounded by the uncertainty principle. Of course, one could claim that the quantum theory is incomplete, and that the knowledge provided to us by the quantum state doesn't really describe *everything* about the "true" state of the system. This mostly becomes an issue of the interpretation of the state and moreover all of quantum theory. We will return to these sort of concerns in Chapter 4.

⁴Remember, no physical normalizable state can have even well-defined position or momentum. But this is beyond the point being made here.

2.3 The Energy-Time Uncertainty Principle

Another pillar of deceit in the popular science community—and even in “proper” physics communities—is the energy-time uncertainty principle. Whenever a physicist invokes the energy-time uncertainty principle, keep a hand on your wallet.⁵ The old version takes the form

$$\Delta E \Delta t \gtrsim \hbar \quad (2.6)$$

and the more modern relation sometimes takes the form

$$\Delta E \Delta t \geq \frac{\hbar}{2}. \quad (2.7)$$

As someone who is now deeply informed about the true modern uncertainty relation, this should already bother you. We could replace this ΔE with the standard deviation of the Hamiltonian $\Delta_{\psi}H$, but what about Δt ? We haven’t come across a “time operator”. Moreover, it can be shown that there is no such observable corresponding to the “time” of a particle, whatever that might mean. Obviously this relation can’t be taken on face value as some special case of Equation 2.3.

Some physicists claim the the relation implies it requires some finite time interval Δt to measure the energy of a system with resolution ΔE . Don’t listen to them. This is complete hogwash and was proven wrong experimentally by Aharonov and Bohm [16]. Other physicists will try to feed you some line about how this relation allows you to violate conservation of energy by amounts ΔE for short periods of time Δt , particularly in the context of the so-called “virtual particles” of particle physics. Again, don’t believe it! Energy is *always* conserved.⁶

⁵I should credit Griffiths [7, p. 56] for this advice, but I vaguely remember Seth Lloyd expressing the same sentiment to me first.

⁶Well, as far as we know. For some excellent guidance on what actually goes on in these circumstances see Section 2.2 of Griffiths’s book on elementary particles [7]. In particular, read page 65, as well as the

So, what can be said about this relation? Well, the energy of a state is related to its frequency.⁷ In order to measure the frequency of something, it needs to stick around long enough for you to see a few cycles. We might then imagine that the uncertainty in the energy of some state is related to the lifetime of that state. However, in quantum mechanics, particles live forever, so we need to be careful about how we define such a state lifetime. Sure enough, by combining the Robertson uncertainty relation with the Ehrenfest theorem⁸

$$\frac{d}{dt} \langle A \rangle_\psi = \frac{1}{i\hbar} \langle [A, H] \rangle_\psi + \left\langle \frac{\partial A}{\partial t} \right\rangle_\psi \quad (2.8)$$

we find that for any observable A that does not explicitly depend on time (so $\langle \frac{\partial A}{\partial t} \rangle = 0$), we have

$$\Delta H \frac{\Delta A}{\frac{d\langle A \rangle}{dt}} \geq \frac{\hbar}{2} \quad (2.9)$$

where everything in the left-hand side is understood to be state dependent. If we interpret the second term in this equation as our Δt , we recover Equation 2.7.⁹ This Δt roughly represents the amount of time it takes the average value of any (not explicitly time-dependent) observable to change by a standard deviation. This isn't exactly correct since both the standard deviation and rate of change of the expectation value can themselves depend on time; however, it's not an altogether terrible way of denoting the "lifetime" of a state as measured by some observable.

This is not the only way we can make sense of the energy-time uncertainty principle. For one, we can derive relations that make use of better definitions of the "lifetime" of the state. We can also derive relations between the duration of some temporary perturbation

footnotes on that page. You should also look over Problem 1.2.

⁷See Equation 1.23.

⁸You have probably noticed that this is just the result of taking the expectation value of the Heisenberg equation of motion, Equation 1.26. While this is true, the Ehrenfest theorem applies equally as well to expectation values of operators taken in the Schrödinger picture. Its proof is very similar to how one derives the Heisenberg equation.

⁹Of course, such a choice of Δt is dependent on the choice of the observable A .

of a system and the size of the energy transitions that the perturbation allows. Yet another derivation leads to a relation useful for analyzing quantum clocks in which a Δt is defined as the amount of time it takes a non-stationary state to evolve into an orthogonal state. These and other approaches are surveyed by Busch [17].

Perhaps the best conclusion to this entire chapter is found in the words of Griffiths [1, p. 118], “But the uncertainty principle is extraordinarily robust: It can be misused without leading to seriously incorrect results, and as a consequence physicists are in the habit of applying it rather carelessly.” Sometimes, its misuse can even lead to deep valid insights that are later proved more formally. So we shouldn’t be *too* hard on people who misuse it. It’s just important to acknowledge its misuse, because sometimes the results are absurd.

Chapter 3

Mixed States

I haven't been entirely up front in my presentation of the postulates of quantum theory. As presented in Chapter 1, these postulates only apply to so-called "pure states". A quantum state *might* be in some pure state, but it might also be in some statistical mixture of pure states. Take the following for example: Say we have some system represented by the two-dimensional complex Hilbert space \mathbb{C}^2 . Let us say $|0\rangle$ and $|1\rangle$ form an orthonormal basis for this space. Consider now the observable σ_z whose matrix in this basis is given by the Pauli z matrix

$$\sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \quad (3.1)$$

If the state of the system is $|0\rangle$ (or $|1\rangle$), a measurement of this observable will return +1 (or -1 respectively) 100% of the time. Now consider the observable σ_x represented in this basis by the Pauli x matrix

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}. \quad (3.2)$$

If the state of the system is $|0\rangle$ or $|1\rangle$, a measurement of this observable will return +1 50%

of the time and -1 50% of the time.

Now let's say the system is in a 50-50 statistical mixture of being in the pure state $|0\rangle$ and the pure state $|1\rangle$. Roughly speaking, this means that half the time it is in state $|0\rangle$ and the other half it is in state $|1\rangle$. If we measure σ_z we get $+1$ or -1 with a 50% chance for each. Likewise if we measure σ_x we will get $+1$ or -1 with a 50% chance for each.

Make sure you see why. Having only learned about pure states, one might assume that we could represent such a state by the pure state $|+\rangle = \frac{1}{\sqrt{2}}|0\rangle + \frac{1}{\sqrt{2}}|1\rangle$ which is a 50-50 *superposition* of $|0\rangle$ and $|1\rangle$. Although this would yield the same statistics for σ_z measurements, the flaw is that a σ_x measurement of this state will *always* return $+1$. (You can carry out a generalized form of this argument to show that such mixed states cannot be represented by *any* pure state ket.)

Of course, we could continue to deal with these states in the way that we just did, i.e. $x_1\%$ of the time the system is in state $|\psi_1\rangle$, and this is what happens; $x_2\%$ of the time the system is in state $|\psi_2\rangle$, and this is what happens; and so on. However, there is a more satisfying way to deal with these mixed states, which is also particularly useful for dealing with composite systems. In this chapter, I will discuss this more useful approach.

3.1 Density Operators

The conventional way of dealing with general states that may be mixed is by using what we call “density operators”. A density operator ρ is a positive-semidefinite Hermitian operator of trace 1 that acts on the state space (the Hilbert space of the system). Positive-semidefinite means that for any vector $|v\rangle$, we have $\langle v|\rho|v\rangle \geq 0$. This is equivalent to saying that all of the eigenvalues of ρ are real and ≥ 0 . Thus, all positive-semidefinite operators are necessarily Hermitian. Trace 1 means that the trace of the operator is one, $\text{tr}(\rho) = 1$.

The trace of an operator is defined as the sum of the diagonal elements of its matrix form

in some orthonormal basis:

$$\text{tr}(\rho) = \sum_i \rho_{ii}. \quad (3.3)$$

This function can be shown to be linear and to possess the property

$$\text{tr}(AB) = \text{tr}(BA) \quad (3.4)$$

for two operators A and B . It follows from this that it doesn't matter what orthonormal basis we choose to represent our operator in when summing up the diagonal matrix elements because a change in orthonormal basis reduces to

$$\text{tr}(U\rho U^\dagger) = \text{tr}(U(\rho U^\dagger)) = \text{tr}((\rho U^\dagger)U) = \text{tr}(\rho(U^\dagger U)) = \text{tr}(\rho) \quad (3.5)$$

for some unitary operator U . This logic also shows that for any diagonalizable operator, such as a Hermitian or unitary operator, the trace is equal to the sum of the eigenvalues of the operator (summing over n copies of each n -fold degenerate eigenvalue). In fact, this can be proven by using triangularization for *any* finite dimensional operator.¹

With the mathematical details out of the way, we can turn back to the physics. The density operator for a pure state² $|\psi\rangle$ is given by $\rho = |\psi\rangle\langle\psi|$. The density operator for a state in a statistical mixture of pure states $|\psi_i\rangle$ with corresponding probabilities p_i is given by

$$\rho = \sum_i p_i |\psi_i\rangle\langle\psi_i|. \quad (3.6)$$

These $|\psi_i\rangle$ need not necessarily be orthogonal. However, due to the spectral theorem, we can

¹This fact generalizes to infinite dimensions as well (where it is called Lidksii's theorem), with the caveat that it applies to so-called "trace class" operators, i.e. operators for which the trace can be defined. The definition of the trace becomes a bit trickier in infinite dimensions, but this technicality is beyond the scope of this paper.

²It follows directly from the way we define the density operator of a pure state that multiplying the state ket by a global phase does not change the corresponding density operator. In many ways, the density operator is a better representation of a pure state than the state vector ket.

decompose any density operator ρ into the form of Equation 3.6 for some (perhaps different) set of $|\psi_i\rangle$ that *are* orthogonal. In this case, the p_i become the eigenvalues and can still be interpreted as probabilities as we are guaranteed they will be ≥ 0 and sum to 1. (Note: This shows an example of how a single mixed state can be decomposed into different sets of pure states and probabilities. It turns out that all mixed states can be decomposed in many ways. This implies that a single mixed state can be interpreted as many different statistical ensembles. This is a very important fact that we will come back to in later chapters.)

The matrix form of a density operator in a given basis is called a “density matrix”. The diagonal elements of such a matrix are called the “populations” and correspond to the probabilities of the state collapsing onto the corresponding basis states after some measurement. The off-diagonal elements of a density matrix are called the “coherences” and contain information regarding how close the state is to being a pure state. If a density matrix has more than one non-zero diagonal element and all of the off-diagonal elements are zero, then the corresponding state is clearly a mixed state. For pure states ρ we have, $\rho = \rho^2$.

3.2 Measurements

By applying our old measurement postulate to the density matrix formalism, one can come up with the following rules. Let’s say we have some observable O whose spectral decomposition is given by Equation 1.9, and our system is in the state ρ . Then a measurement³ of O will return λ_i with probability $\text{tr}(\rho P_{\lambda_i})$, and the system will collapse onto the state $\frac{P_{\lambda_i}\rho P_{\lambda_i}}{\text{tr}(\rho P_{\lambda_i})}$.

If we don’t actually look at the measurement result, it could be anything: the new state could be in any one of the many eigenstates with certain probabilities. Summing over the corresponding density operators multiplied by the corresponding probabilities, the density

³Again I am only considering projective measurements. For more general types of measurements, see Nielsen and Chuang [3].

operator that represents this statistical ensemble is given by

$$\rho' = \sum_i P_{\lambda_i} \rho P_{\lambda_i}. \quad (3.7)$$

Furthermore, it can be shown that the expectation value of the measurement is given by

$$\langle O \rangle_\rho = \text{tr}(\rho O). \quad (3.8)$$

3.3 Evolution

Recalling Equation 1.16 and applying it to Equation 3.6, we have that density operators evolve (in the Schrödinger picture) according to

$$\rho(t) = U(t, t_0) \rho(t_0) U^\dagger(t, t_0). \quad (3.9)$$

Equivalently, we can reformulate the Schrödinger equation (Equation 1.15) for density operators. In which case, we come up with the Liouville–von Neumann equation

$$i\hbar \frac{\partial}{\partial t} \rho(t) = [H, \rho(t)]. \quad (3.10)$$

The Heisenberg picture is the same as before. In the words of a famous shrimper, “That’s all I have to say about that.”⁴

⁴Well, maybe not. When you saw Equation 3.10 you probably realized it looks like the Heisenberg equation of motion (Equation 1.26), but without the often-unimportant final term. Upon closer inspection, however, you’ll see that there is a sign difference, or equivalently, that the commutator is taken in the opposite order. How can you prevent yourself from mixing these two equations up? Simple. First, just remember that there *is* a difference. Then, remember that the equation for states (density operators) is the one that looks most like Schrödinger’s equation, i.e. that the H is to the *left* of the state. This sign difference reflects the fact that the Heisenberg picture observables evolve in the opposite way that Schrödinger picture density operators evolve. Now, compare Equation 1.25 to Equation 3.9. Either we hold the observables fixed and rotate the states “forwards” in time (Schrödinger picture), or we hold the states fixed and rotate the observables “backwards” in time. Either way, the relative rotation between the states and observables is the

3.4 Composite Systems

As one would expect, the density operator corresponding to the state of a composite system whose subsystem state spaces are given by \mathcal{H}_A and \mathcal{H}_B is given by a positive-semidefinite trace 1 density operator that acts on the joint state space $\mathcal{H}_A \otimes \mathcal{H}_B$. All such density operators can be written as some affine⁵ combination of simple products states

$$\rho = \sum_{i,j} \alpha_{i,j} \tau_i \otimes \nu_j \quad (3.11)$$

where τ_i are density operators of the first subsystem and ν_j are density operators of the second subsystem.

The simple product state $\rho = \rho_A \otimes \rho_B$ represents a state where the first subsystem is in the state ρ_A and the second subsystem is in state ρ_B .⁶ This inspires us to define an operation called the “partial trace” with respect to subsystem B on tensor products of linear operators by

$$\text{tr}_B(L_A \otimes L_B) = L_A \text{tr}(L_B) \quad (3.12)$$

which returns a linear operator that acts on \mathcal{H}_A . Likewise we define the partial trace with

same. Hence the equivalence of the two pictures. (On a different note, you should now realize that the joke from the footnote on page 22 is just as funny in the Schrödinger picture as it is in the Heisenberg picture. I always want to somehow work in the idea that Zog and Thug work as operators of some kind, but it’s very easy to ruin a good joke by overworking it.) There is yet a third picture that I haven’t mentioned called the “interaction” (or “Dirac”) picture. In this picture, the states evolve by rotating forward in time *a little bit*, and the operators rotate backwards in time *a different little bit*, such that the net effect is the same total relative rotation between the two. The Hamiltonian is split into two parts: $H = H_0 + H_1$. The first part H_0 is typically time-independent and simple enough to solve exactly. The second part H_1 is often a time-dependent potential $V(t)$. States in the interaction picture obey a modified Schrödinger or Liouville–von Neumann equation (where H is replaced by H_1), and operators in the interaction picture obey a modified Heisenberg equation of motion (here H is replaced by H_0). For a treatment of this topic in the context of time-dependent potentials, see Sakurai [4, pp. 336–339]. (I leave it to you to work out the details of how—or if—Zog and Thug fit into the interaction picture.)

⁵An affine combination is a linear combination where the coefficients sum up to 1. Note the coefficients are still allowed to be negative or greater than 1. When I use the term affine I also imply that the coefficients are real.

⁶For pure simple product states $|\psi\rangle = |\phi_A\rangle \otimes |\phi_B\rangle$, we have $\rho_A = |\phi_A\rangle\langle\phi_A|$ and $\rho_B = |\phi_B\rangle\langle\phi_B|$. Sure enough, $\rho = |\psi\rangle\langle\psi| = (|\phi_A\rangle \otimes |\phi_B\rangle)(\langle\phi_A| \otimes \langle\phi_B|) = |\phi_A\rangle\langle\phi_A| \otimes |\phi_B\rangle\langle\phi_B| = \rho_A \otimes \rho_B$.

respect to subsystem A by

$$\text{tr}_A(L_A \otimes L_B) = \text{tr}(L_A)L_B \quad (3.13)$$

which returns a linear operator that acts on \mathcal{H}_B . We extend the definition of the partial trace(s) to arbitrary linear operators that act on $\mathcal{H}_A \otimes \mathcal{H}_B$ by linearity.⁷ If we return to our simple product state $\rho = \rho_A \otimes \rho_B$, we find that $\text{tr}_B(\rho) = \rho_A \text{tr}(\rho_B) = \rho_A$ and likewise that $\text{tr}_A(\rho) = \rho_B$. As we can see, tracing out⁸ a subsystem of a simple product state returns the state of the other subsystem. Is this true for general states that aren't simple product states?

Let ρ be some general density operator given by an affine combination of simple product states. Let O be some observable on subsystem A , with spectral decomposition given by Equation 1.9. Then, one can prove the following relations:

$$\text{tr}_B((P_{\lambda_i} \otimes I)\rho(P_{\lambda_i} \otimes I)) = P_{\lambda_i} \text{tr}_B(\rho)P_{\lambda_i} \quad (3.14)$$

$$\text{tr}(\rho(P_{\lambda_i} \otimes I)) = \text{tr}(\text{tr}_B(\rho)P_{\lambda_i}) \quad (3.15)$$

$$\text{tr}(\rho(O \otimes I)) = \text{tr}(\text{tr}_B(\rho)O) \quad (3.16)$$

and the similar relations where the roles of A and B are switched. From this and the measurement postulate we can conclude that the tracing out a subsystem of a general density operator *does* return the proper “reduced density operator” for the other system.⁹ We use this term “reduced density operator”¹⁰ to emphasize that this isn’t necessarily *the* state of the other subsystem because in general $\rho \neq \text{tr}_B(\rho) \otimes \text{tr}_A(\rho)$. However, a local observer will

⁷Remember that all such linear operators can be represented as linear combinations of simple tensor products of local linear operators.

⁸We refer to taking the partial trace with respect to a subsystem as “tracing out” that subsystem.

⁹This reflects the mathematical fact that, in quantum theory, a state is completely determined by the results it yields for all possible measurements. Whether or not this is true physically is an interesting philosophical question.

¹⁰Sometimes the term “marginal state” is used.

find the measurement statistics of any local measurement to be the same as if his subsystem was in fact in the state given by the reduced density operator.

A perfect example of this subtle distinction is that there exist bipartite pure states whose reduced density operators are mixed. (These states turn out to be quite special, and are the subject of the next chapter.) This fact poses an interesting question. If the subsystem of some composite system in a pure state can be effectively in a mixed state, then does it really make sense to consider mixtures as simply statistical ensembles? Should we consider mixed states as a fundamental state that a *single* particle can be in? This is more of a philosophical question than a question of physics, but an interesting mathematical fact related to this question is that it is always possible to represent any mixed state of any system as the reduced density operator of some larger pure (entangled) state of a larger, composite system where we tack on some ancillary Hilbert space. This is called the “purification” of the quantum state.

Members of the Church of the Larger Hilbert Space believe that all physical instances of mixed states are necessarily just the reduced density operators of some larger pure state for which the other subsystems aren’t being considered. Members of the Church of the Smaller Hilbert Space explicitly reject this belief (see Leifer [18]). Of course, as with any church, these churches contain a number of other tenets, and it is not necessarily true that someone is a member of one or the other.

Chapter 4

Entanglement

In this chapter, I explore some of the basic details surrounding what are known as “entangled states”. These states have many interesting physical, philosophical, and mathematical consequences. They are at the center of much modern study in quantum foundations, quantum information theory, quantum computation, and other fields. Accordingly, I cannot hope to even graze the surface of the topic of entanglement. With that said, this chapter should provide you with enough to understand entanglement in other literature.

4.1 Pure Entangled States

As I mentioned in Section 1.4, pure states are called entangled when they cannot be written in the form of a simple product state such as $|\psi\rangle = |\phi_A\rangle \otimes |\phi_B\rangle$. States that are not entangled (and thus *can* be written as a simple product in the pure state case) are called “separable”.

By the definition of the tensor product of vector spaces, any pure state of a composite system (including entangled states) can be written as a linear combination of product states. Furthermore, there is a result in linear algebra known as the Schmidt decomposition which guarantees that for any state $|\psi\rangle$ in $\mathcal{H}_A \otimes \mathcal{H}_B$, there exist a basis $|u_i\rangle$ for \mathcal{H}_A and a basis $|v_i\rangle$

for \mathcal{H}_B such that

$$|\psi\rangle = \sum_i \alpha_i |u_i\rangle \otimes |v_i\rangle \quad (4.1)$$

where the α_i 's are real numbers ≥ 0 , $\sum_i \alpha_i^2 = 1$, and are called “Schmidt coefficients”. In general, the two Hilbert spaces may not be of the same dimension. In such a case, the sum ranges over all the dimensions of the smaller space, and some of the basis vectors of the larger space aren't used. A given state $|\psi\rangle$ has only one set of Schmidt coefficients. It follows from this that entangled states are states that have two or more non-zero Schmidt coefficients.

If we consider the reduced density operators of $|\psi\rangle$ we see that

$$\text{tr}_A(|\psi\rangle\langle\psi|) = \sum_i \alpha_i^2 |u_i\rangle\langle u_i| \quad (4.2)$$

$$\text{tr}_B(|\psi\rangle\langle\psi|) = \sum_i \alpha_i^2 |v_i\rangle\langle v_i| \quad (4.3)$$

where the eigenvalues of either reduced density operator are given by α_i^2 . This implies that the both¹ of the reduced density operators of entangled states are mixed. Conversely, it is true that all pure states that have mixed marginals are necessarily entangled.² Thus, pure entangled states are precisely the states for which $|\psi\rangle\langle\psi| \neq \text{tr}_A(|\psi\rangle\langle\psi|) \otimes \text{tr}_B(|\psi\rangle\langle\psi|)$. These states fly in the face of the reductionist idea that a system is merely the sum of its parts.

4.2 General Entanglement and the Minus Sign

As shown by Equation 3.11, all general composite states ρ can be written as an affine combination of general simple product states. General separable states are the states that can be written as a *convex*³ combination of general simple product states. Entangled states

¹There are no pure composite states that have one pure marginal and one mixed marginal.

²This is a good place to introduce the mathematical term “iff”. We say that a pure state is entangled iff (or “if and only if”) its marginals are mixed.

³A convex combination is an affine combination with the further restriction that all of the coefficients are ≥ 0 . This also implies that the coefficients are ≤ 1 , and thus that they can be interpreted as probabilities.

then, are the states that cannot be written as a convex combination of product states. For pure states, this reduces to the condition from the previous section. Although they cannot be written as convex combinations of product states, they can still be written as *affine* combinations of product states.

A brief digression: In light of their affine decompositions, we could interpret entangled states as mixtures of product states that include negative probabilities (and subsequently, probabilities greater than 1). Such “probabilities” would still be guaranteed to sum to one. This is an interesting interpretation, but I am not sure if anything deeper can be said about it.⁴ This is related to the “Minus-Sign Test” proposed by Aaronson [19]. Aaronson believes that mentioning the minus signs in quantum mechanics is the sole thing a popularization of quantum mechanics needs to do to sufficiently express the theory and avoid making misleading statements, particularly about superposition. Of course, Aaronson was in part referring to the relative phase differences that can be present in pure state superpositions and ultimately allow for interference (as I will discuss in Chapter 5). In general, these phases can be complex, not just negative. In case you think this suggests that the Minus-Sign Test is an oversimplification, even single partite superpositions can be reformulated in terms of negative probabilities by appealing to Wigner quasi-probability distributions.⁵ Furthermore, there is an entire reformulation of quantum mechanics in terms of real Hilbert spaces instead of complex ones.⁶ Seen in these ways, Aaronson’s sentiment is completely justifiable to any standard of rigor. One might say that minus signs are *the* thing that distinguish quantum theory from classical probability theory, and that complex numbers are just tools for mere

⁴Dirac and Feynman were both known to have independently endorsed the applications of negative probabilities to quantum mechanics.

⁵The Wigner quasi-probability distribution is a quantum generalization of the classical notion of x , p phase space. It’s another way of representing states. I don’t know if the negative probabilities it introduces have anything to do with the negative probabilities arising from decomposing entangled states into affine combinations, but I suspect they do. I once asked a grad student at IQC about it. He replied with a blank stare. He must not have known either.

⁶See Stueckelberg [20]. Stueckelberg is just the tip of an iceberg that I omit, but one has to show respect for people who publish in Helvetica. (I kid.)

mortals like ourselves to make sense of quantum theory in terms of classical wave theory intuitions.

To return to the non-reductive nature of entangled states: Note, there exist general *separable* states that cannot be written as a simple product state, i.e. $\rho \neq \text{tr}_B(\rho) \otimes \text{tr}_A(\rho)$. These correspond to states with classical correlations. WRONG! SAY SOMETHING ABOUT QUANTUM DISCORD! At first these states also seem to fly in the face of reductionists, but since they are necessarily mixtures (convex combinations) of simple product states, a reductionist would probably claim that the system is *really* in *one* of the product states with certain probabilities. As I will discuss in the next two sections, this is a tough claim to make concerning entangled states: entangled states resist reductive analysis.

Contrary to classically correlated states, many entangled states exhibit correlations *stronger* than those permitted by classical physics. Typically, if a local measurement is made on a subsystem in some state entangled with another subsystem, the measurement results will imply something about the other subsystem as well. These entangled correlations tend to be present for a number of different observables for a single entangled state: rewriting entangled states in other bases doesn't change the fact that they are entangled. This fact will reappear in the next sections and latter in this paper. Indeed, it is at the heart of my experiment.

4.3 The EPR Thought Experiment

Perhaps the most famous event surrounding entangled states in the development of quantum mechanics was a thought experiment proposed by Einstein, Podolsky, and Rosen (EPR) [21]. In their paper, the authors proposed a thought experiment they claimed was proof that quantum mechanics was necessarily an incomplete theory. Ironically, they used entanglement, a special feature of quantum theory, to show that the theory itself was insufficient to account for what they considered to be all reality.

Their argument was roughly as follows: First, they claimed that for any complete physical theory, “*every element of the physical reality must have a counterpart in the physical theory.*”⁷ Then they went on to provide a sufficient condition for the existence of an element of physical reality with the statement:

If, without in any way disturbing a system, we can predict with certainty (i.e. with probability equal to unity) the value of a physical quantity, then there exists an element of physical reality corresponding to this physical quantity.

Next, they remarked on the fact that there exist quantum mechanical states of two particles for which the particles’ positions and momenta are entangled, in that by measuring either the position or momentum of one of these particles, one can predict with absolute certainty the corresponding value for the other particle. They proposed that we consider a system of two particles that are allowed to interact for some brief time, in which they become entangled, “after which time we suppose that there is no longer any interaction”.⁸ Then, looking at the particle we still have, we could choose to measure its position or momentum. By doing so, we could predict *either* the position *or* the momentum of the second particle *without disturbing it*. Thus, according to their argument, *both* the position and momentum of the second particle *must* be elements of physical reality. However, quantum mechanics does not contain states of particles that have a well-defined position *and* momentum. Such states would violate the uncertainty principle. So, Einstein, Podolsky, and Rosen concluded that the quantum mechanical description of reality must be incomplete.

Although they admitted that their conclusion would not survive a more restrictive definition of reality where “two or more physical quantities can be regarded as simultaneous

⁷The emphasis in this and all quotes in this paper are in the originals.

⁸The authors weren’t clear about how we could enforce this assumption. In retrospect, we imagine that one particle is sent far, far away from the other. It is assumed they cannot interact because they are far apart: any interaction would violate the universal speed limit of interactions c from relativity. According to quantum mechanics, this separation would have no effect on the entanglement. These and other subtle assumptions that are made throughout the EPR paper will be the topic of the next section.

elements of reality *only when they can be simultaneously measured or predicted*”, they claimed that this definition was absurd. Such a definition would insist that since we have to *choose* whether to (1) measure the position of the second particle without disturbing it *or* (2) measure the momentum of the second particle without disturbing it,⁹ then the argument does not show that the position and momentum of the second particle are both *simultaneous* elements of reality. The authors remarked about such insistence that

This makes the reality of [the position and momentum of the second particle] depend upon the process of measurement carried out on the first system, which does not disturb the second system in any way. No reasonable definition of reality could be expected to permit this.

4.4 The Bell and CHSH Inequalities

A simplified version of the EPR thought experiment involving spin- $\frac{1}{2}$ particles was proposed by Bohm [22, p. 614]. In this proposal, one considers an entangled pair of spin- $\frac{1}{2}$ particles. The state space of each particle is the two-dimensional complex Hilbert space \mathbb{C}^2 . Thus, the joint state space is the four-dimensional Hilbert space $\mathbb{C}^2 \otimes \mathbb{C}^2 \cong \mathbb{C}^4$. One then considers the observables represented by passing the particles through local Stern-Gerlach analyzers that are oriented at certain angles. These observables are characterized by the orientation angle of the analyzer. Such analyzers return a measurement value of +1 or -1 for spin- $\frac{1}{2}$ particles.¹⁰ Pairs of angles for Stern-Gerlach orientations can be chosen that are represented by observables analogous to the position and momentum of the EPR proposal. Likewise, there exist entangled states that correlate these observables between two particles. The state

⁹We can't measure both simultaneously because that would require simultaneous measurements of the position and momentum of the first particle (which don't commute).

¹⁰For those with a good memory, yes, the Pauli matrices from Chapter 3 are examples of Stern-Gerlach observables.

that Bohm's proposal concerns itself with is the so called “singlet state”

$$|\Psi^-\rangle = \frac{1}{\sqrt{2}}(|0\rangle \otimes |1\rangle - |1\rangle \otimes |0\rangle) \quad (4.4)$$

for some $|0\rangle$ and $|1\rangle$ that form an orthonormal basis of the two local Hilbert spaces. This state has the interesting property of taking the same form in any basis (provided we choose the *same* local basis for each subsystem); it doesn't matter what we choose for $|0\rangle$ and $|1\rangle$.

In this context, Bell [23] realized that the assumptions made in the EPR paper not only implied that quantum mechanics was incomplete, but the assumptions were actually *incompatible* with quantum mechanics. Bell reduced the assumptions made by the EPR paper to (1) *locality*, the assumption that actions at some location cannot influence systems at distant positions arbitrarily quickly or instantaneously (motivated by relativity); and (2) *hidden variables*, the assumption that some “real” description of reality exists which can be modeled by some yet unknown set of quantities. However, this second assumption isn't truly part of the EPR paper. As the authors note at the end of the paper

While we have thus shown that the wave function does not provide a complete description of the physical reality, we left open the question of whether or not such a description exists. We believe, however, that such a theory is possible.

Later work done by Stapp [24, 25] and Eberhard [26, 27, 28], among others, was able to re-derive Bell's results (typically referred to as “Bell's theorem”) by replacing the hidden variable assumption with the true assumption the EPR paper makes: *counterfactual definiteness*, which is the assumption that one can speak meaningfully about the definiteness of the results of measurements that have not been performed. The assumptions of locality and counterfactual definiteness taken together are known as “local realism”.

Let $A(\theta_A)$ be the observable corresponding to a Stern-Gerlach measurement (oriented at angle θ_A) of the first particle, and let $B(\theta_B)$ be the observable corresponding to a Stern-

Gerlach measurement (oriented at angle θ_B) of the second particle. What Bell realized was that the quantum correlation of such measurements of the local subsystems of a singlet is given by

$$C(A(\theta_A), B(\theta_B)) = \langle A(\theta_A) \otimes B(\theta_B) \rangle_{\Psi^-} = -\cos(\theta_A - \theta_B) \quad (4.5)$$

i.e. it is given by a sinusoid that varies with the relative angle between the two analyzers. On the other hand, Bell showed that the natural local hidden variable interpretation of the singlet state gives this correlation a *linear* shape

$$C(A(\theta_A), B(\theta_B)) = -1 + \frac{2}{\pi}(\theta_A - \theta_B). \quad (4.6)$$

These two different graphs are plotted together in Figure 4.1.¹¹ In any theoretical context, we define the correlation $C(A, B)$ for properties A and B that can each take the values $+1$ or -1 to be the average value of the product of the measurement results obtained from measuring A on the first subsystem and B on the second subsystem.

Bell went on to show that for *any* local hidden variables theory, the correlations of the observables for three angles θ_1 , θ_2 , and θ_3 must obey the inequality

$$1 + C(\theta_2, \theta_3) \geq |C(\theta_1, \theta_2) - C(\theta_1, \theta_3)| \quad (4.7)$$

commonly referred to as “Bell’s inequality”, where I have used the angles themselves to denote the observables representing a Stern-Gerlach measurement oriented at that angle.

¹¹So, the next time you see some kid in high school sketching sinusoids for the first time by plotting the values at integer multiples of $\frac{\pi}{2}$ and just connecting the dots with straight lines, tell them “No, no, no! You have to curve the lines!” When they respond “What difference does it make?”, you tell them “The difference between classical intuition and quantum physics. That’s the difference it makes!” I had this same dialogue recently with my fiancée’s little brother. A day or so later I was deriving the equations of motion of the simple double pendulum to put together a final project for a math class. You end up with a page or so full of simple, but very messy equations. He saw this mathematical mess, and took a picture of it with his phone. I gave him a puzzled look, and he said “My teacher wants to know what you do.” I said, “Oh, well this is far from what I do.”

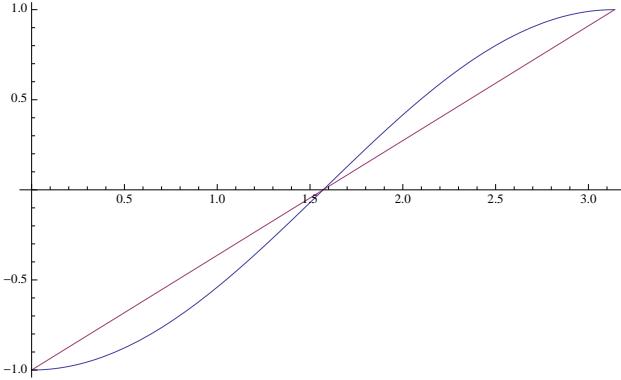


Figure 4.1: A plot of the Bell correlations of the spin- $\frac{1}{2}$ singlet state. The vertical axis represents the quantum correlation $C(A(\theta_A), B(\theta_B)) = \langle A(\theta_A) \otimes B(\theta_B) \rangle_{\Psi^-}$. The horizontal axis is the difference in the angles $\theta_A - \theta_B$ in radians. The blue plot shows the quantum mechanical prediction. The purple plot shows the predictions of the natural hidden variables approach to the singlet state.

On the contrary, the singlet state violates this inequality for certain choices of angles. Thus, the local realism assumptions are incompatible with quantum mechanics! Whether or not one can interpret the world under these assumptions now became a question that could be tested experimentally.

Clauser, Horne, Shimony, and Holt [29] went on to come up with a more general form of Bell's theorem. For any bipartite quantum system where A_1 and A_2 are observables on the first subsystem with possible outcomes $+1$ and -1 , and likewise B_1 and B_2 are observables on the second subsystem with possible outcomes $+1$ and -1 , local-realistic theories must obey the inequality

$$-2 \leq C(A_1, B_1) + C(A_1, B_2) + C(A_2, B_1) - C(A_2, B_2) \leq 2. \quad (4.8)$$

The quantity in the center is sometimes referred to as the “CHSH parameter” denoted by S . Next I will show a rough proof of this inequality adapted from Nielsen and Chuang [3,

p. 115].¹²

Assume the *real* state of the system is such that with some probability $p(a_1, a_2, b_1, b_2)$ the property A_1 *really* has the value a_1 , the property A_2 *really* has the value a_2 , the property B_1 *really* has the value b_1 , and the property B_2 *really* has the value b_2 . All of these numbers are either $+1$ or -1 . The probability might have to do with experimental noise, the preparation of the state, or some intrinsic randomness of reality. Since expectation values are linear (even when the properties are not statistically independent), we have

$$S = \sum_{a_1, a_2, b_1, b_2} p(a_1, a_2, b_1, b_2)(a_1 b_1 + a_1 b_2 + a_2 b_1 - a_2 b_2) \quad (4.9)$$

$$= \sum_{a_1, a_2, b_1, b_2} p(a_1, a_2, b_1, b_2)\{(a_1 + a_2)b_1 + (a_1 - a_2)b_2\}. \quad (4.10)$$

For any individual choice of the numbers, either $a_1 + a_2 = 0$ or $a_1 - a_2 = 0$ since $a_1 = \pm 1$ and $a_2 = \pm 1$. Whichever one is 0, the other will necessarily be ± 2 . Furthermore, $b_1 = \pm 1$ and $b_2 = \pm 1$. So for any choice of the numbers, $(a_1 + a_2)b_1 + (a_1 - a_2)b_2 = \pm 2$. Thus, the *most* S could be is

$$S_{\max} = \sum_{a_1, a_2, b_1, b_2} p(a_1, a_2, b_1, b_2) \times 2 \quad (4.11)$$

$$= 2 \times \sum_{a_1, a_2, b_1, b_2} p(a_1, a_2, b_1, b_2) \quad (4.12)$$

$$= 2 \quad (4.13)$$

¹²Bell's theorem was the toughest part of quantum theory for me to swallow when I was first learning it. Before having seen a proof, I would try to think up some local hidden variable theory, fail, and continue to struggle on. I show this proof to ensure you don't suffer the same fate.

and the *least* S could be is

$$S_{\min} = \sum_{a_1, a_2, b_1, b_2} p(a_1, a_2, b_1, b_2) \times -2 \quad (4.14)$$

$$= -2 \times \sum_{a_1, a_2, b_1, b_2} p(a_1, a_2, b_1, b_2) \quad (4.15)$$

$$= -2. \quad (4.16)$$

So, we have $-2 \leq S \leq 2$.

Now I encourage you to consider a quantum system in the singlet state, let $A_1 = \sigma_z$, $A_2 = \sigma_x$, $B_1 = \frac{1}{\sqrt{2}}(-\sigma_z - \sigma_x)$, and $B_2 = \frac{1}{\sqrt{2}}(\sigma_z - \sigma_x)$, where σ_z and σ_x are the observables given in Chapter 3. Evaluate the CHSH parameter for this system and observables, and you'll find that $S = 2\sqrt{2} > 2!$ ¹³ So quantum mechanics is not compatible with local realism! Moreover, this is a real state we can prepare in the lab and these are real observables we can measure. So, the question of whether or not local realist assumptions can be made about the real world can be experimentally tested.

Most experimental tests of Bell's theorem use the above CHSH inequality, the related CH74 inequality derived by Clauser and Horne [30], or some variation of these inequalities. The first such experiment was performed in 1972 by Freedman and Clauser [31]. The results agreed with the quantum mechanical prediction with high statistical accuracy. However, proponents of local realistic interpretations of the world have objected that there are a number of loopholes in this and other experiments that were later performed. For one, they claim that the detector efficiencies are usually not high enough to invoke the necessary “fair-sampling” assumption. They also note that in many experiments, the subsystems are close

¹³That's an exclamation mark for the sentence, not to be interpreted as the factorial of 2—although in this case it doesn't make a mathematical difference. (Of course, now I've made the problem worse by adding a superscript number that looks like an exponent that *does* make a mathematical difference. But I digress.) As a student in freshman physics, I once had the balls to do one of the labs with 5 other partners. When the instructor returned the lab I found “# of partners must be $< 3!$ ” written in red ink. So I wrote $5 < 6 = 3!$ and put it back in her mailbox.

enough to each other and the measurements are made slowly enough that the two systems could communicate with each other in the interim without violating the relativistic speed limit c .

In 1998 Weihs and a team led by Zeilinger performed the first Bell experiment to close the latter loophole, also known as the “locality” loophole (see Weihs et al. [32]). Their experiment violated of the CHSH inequality by more than 30 standard deviations, with data that agreed with quantum predictions. However, they were unable to simultaneously close the former loophole, also known as the “detection” loophole. The first experiment to close the detection loophole was performed in 2001 by Rowe et al. [33] with detection efficiencies well over 90%. However, the two subsystems were very close together and the locality loophole could not be closed. Many other experiments have been carried out, all of which agree with quantum mechanics, but none have been able to simultaneously close the two major loopholes. Most physicists at this point are convinced by the experimental evidence we *do* have. To be unconvinced requires beliefs that the physics of the world is working in a very conspiratorial way. Nevertheless, a loophole-free Bell experiment is still a large goal in the experimental community. It should also be noted that almost all of these experiments have used the Bohm version of the EPR setup. The first EPR-type experiment using the original position and momentum context was performed using single photons in 2004 by Howell, Bennink, Bentley, and Boyd [34]. It should be noted that almost every EPR-type experiment performed to date has used single photons as the “particles”. As discussed by Clauser and Shimony [35], even short pulses of classical light are sufficient.

As it turns out, violating the CHSH inequality is a sufficient condition to determine that one is in possession of an entangled state. But, perhaps surprisingly, there are entangled quantum states that don’t violate Bell-type inequalities such as the CHSH inequality. Nevertheless, an interesting question is: What is the *maximal* violation of the CHSH parameter permitted by quantum theory? The answer turns out to be $2\sqrt{2}$ as shown by Tsirelson [36]

in 1980. This number is known as Tsirelson’s bound. What’s so special about it? The 2 for local realistic theories can be reasoned from intuitive principles. Why should $2\sqrt{2}$ be the quantum bound? The strictly algebraic bound is 4.¹⁴ One can construct generalized probability theories¹⁵ that saturate this bound, as Popescu and Rohrlich [38] did in 1994. So, is there some intuitive reason why Nature should impose Tsirelson’s bound on physical states? This is a huge open question in quantum foundations. A related question is whether there are some intuitive physical principles (analogous to the Laws of Thermodynamics) from which the postulates of quantum theory can be derived. As they stand, the postulates of quantum theory and Tsirelson’s bound both seem extremely ad-hoc. Some work seeks to explicitly kill both birds with the same stone. Much work has been done in this area, and some concepts have been proposed that have partially recovered quantum theory. Likewise, much effort has been put into formulating other theorems like Bell’s that put restrictions on what type of assumptions one can make about the way the world really works, often with an experimental verification proposal. The most famous example is the Kochen-Specker theorem [39]. An extremely recent example is the work of Pusey, Barrett, and Rudolph [40].

Another question a reader might ask: Does Bell’s theorem imply that entangled states can be used to send messages faster than light or, subsequently, travel back in time? The unfortunate answer is no. Bell’s theorem does not imply that quantum mechanics is necessarily nonlocal. Moreover, it can be shown that local operations and measurements on a subsystem cannot change any other subsystem’s marginal state without some traditional transmission of information, even for entangled states. This is known as the “no-communication theorem”

¹⁴The CHSH parameter is a sum of 4 terms each of which cannot be greater than 1 or less than -1.

¹⁵For an introduction to the field of generalized probability theory and a treatment of quantum theory as a special case, see Wilce [37].

or “no-signaling condition”.¹⁶ It would appear that we can’t have our cake and eat it too.¹⁷ In light of the previous paragraph, one might be tempted to postulate that the this restraint imposed by relativity is what enforces Tsirelson’s bound. This is not the case, at least not in any obvious way. Very many generalized probability theories obey the no-signaling condition; when talking about generalized probability theories it is essentially tacitly assumed that one is talking about *non-signaling* theories. So, as Popescu and Rohrlich [38] showed, you can algebraically saturate the CHSH parameter without allowing for composite states that can be used to instantaneously communicate.

4.5 Entanglement as a Resource

I conclude this chapter with a very brief acknowledgment that entangled states can serve as a useful physical resource. One such example is found in quantum teleportation. Due to the no-cloning and no-broadcasting theorems¹⁸, it is impossible to make a perfect copy of any quantum system while leaving the original intact. In spite of this, pre-shared entanglement allows one to teleport a quantum state to a distant party, destroying the original in the process. This does not violate the no-communication theorem, because some small amount of classical information must be sent to complete the process. Entanglement can also be used as a computational resource to create computational algorithms that beat our

¹⁶See A. Peres and D. Terno [41]. Some authors such as Kennedy [42], Peacock and Hepburn [43] claim that these proofs are circular; that the no-signaling condition is built into the assumptions of the composite system postulate. In and of itself, this does not imply that there is *no* way of communicating faster than light. For a more general discussion of this impossibility in the context of quantum field theory, see Eberhard and Ross [44]. Sending particles faster than light is usually presumed to be impossible due to relativity. This is not true however. Faster-than-light (FTL) particles called tachyons could exist without violating relativity, but they would have many strange properties such as imaginary mass (or alternatively, imaginary energy and momentum). They could also travel backwards in time. However, attempts to introduce tachyonic quantum fields with imaginary masses ultimately fail to allow FTL particle travel. One way or another, experimental discovery of FTL phenomena would almost certainly indicate new physics.

¹⁷This is a recurring theme in quantum theory. For example, you can’t know anything about some quantum state without measuring it and thus ruining it. There is much evidence to suggest that nature is a big cosmic joke. Try to laugh when you see its irony.

¹⁸See Wootters and Zurek [45], Dieks [46], and Barnum et al. [47].

best known classical algorithms. Such is the study of quantum computation. These examples and further discussion of entanglement as a resource can be found in Nielsen and Chuang [3]. Such proposals form a large part of modern quantum studies, particularly in the areas of quantum computation and quantum information theory. The most recent work has been concerned with developing applications of entanglement to communication, such as the quantum internet infrastructure proposed by Lloyd, Shaprio, Wong, et al. [48].

In light of all the recent concerns about internet privacy, one might also be interested to know that experimentally demonstrated proposals such as quantum encryption and blind quantum computing exist that are provably secure. LLoyd [49] and others have gone as far as proposing quantum search engines: a “Quoogle” that could never snoop on its own users’ data. Not surprisingly, Google didn’t jump on his bandwagon when LLoyd pitched the idea. Nevertheless, there is still a realistic hope that some day in the distant future our great-great-grandchildren will all be using secure quantum computers, communicating on a quantum internet using quantum encryption. Quantum encryption is already being used in the real world, such as at the FIFA World Cup [50].

Chapter 5

The Double-Slit Experiment

Perhaps the most ubiquitous experiment in quantum mechanics is the so-called “double-slit” experiment. To quote Feynman [51, p. 1-1]:

In this chapter we shall tackle immediately the basic element of the mysterious behavior in its most strange form. We choose to examine a phenomenon which is impossible, *absolutely* impossible, to explain in any classical way, and which has in it the heart of quantum mechanics. In reality, it contains the *only* mystery. We cannot make the mystery go away by “explaining” how it works. We will just *tell* you how it works. In telling you how it works we will have told you about the basic peculiarities of all quantum mechanics.

5.1 The Party Circa 1800

Early in the history of physics, Gassendi proposed that light was made up of particles. This later inspired Newton to come up with his own particulate theory of light from which he derived many common optical phenomena. The particle theory of light was the generally accepted theory in the seventeenth century, although Hooke and Huygens independently

came up with wave theories of light around the same time. Finally, around 1800, Thomas Young performed the double-slit experiment which verified the wave nature of light. Fresnel then went on to come up with his own wave theory, which showed that the polarization of light could only be explained by a transverse wave. This work was expanded upon by Poisson. For these and other reasons, the old particle theory of light was overturned in the nineteenth century.

A sketch of the classical double-slit experiment is shown in Figure 5.1. A picture of the resulting interference pattern in the detection plane (as well as the single-slit diffraction¹ case) is shown in Figure 5.2. The experiment consists of a coherent² source of light that is incident on a barrier with two slits the light can pass through. These figures depict the modern version of the experiment which uses a laser for the light source. Young's original experiment used sunlight as a source. He used an initial single slit to pick out a coherent piece of the sunlight. What he found was that the light on the other side of the double slit fell in bands of light and dark, precisely the type of pattern seen in Figure 5.2. This pattern is easily be explained in terms of waves, but cannot be explained in terms of particles of light (at least not in any classical way).

A rough argument as to how the double-slit interference pattern is described by a wave theory is shown in Figure 5.3. Without repeating the description found in the caption, the argument is roughly as follows: Waves are often solutions to certain linear differential

¹The distinction between interference and diffraction is fuzzy at best. Feynman [52, p. 30-10] says:

No one has ever been able to define the difference between interference and diffraction satisfactorily. It is just a question of usage, and there is no specific, important physical difference between them. The best we can do, roughly speaking, is to say that when there are only a few sources, say two, interfering, then the result is usually called interference, but if there is a large number of them, it seems that the word diffraction is more often used. So, we shall not worry about whether it is interference or diffraction, but continue from where we left off...

²Coherence is a term physicists use to measure how tidy a wave is. Quite literally, it is by definition the property that measures an ability of a wave to interfere. If you are unfamiliar with the term, try not to get too worried about the details.

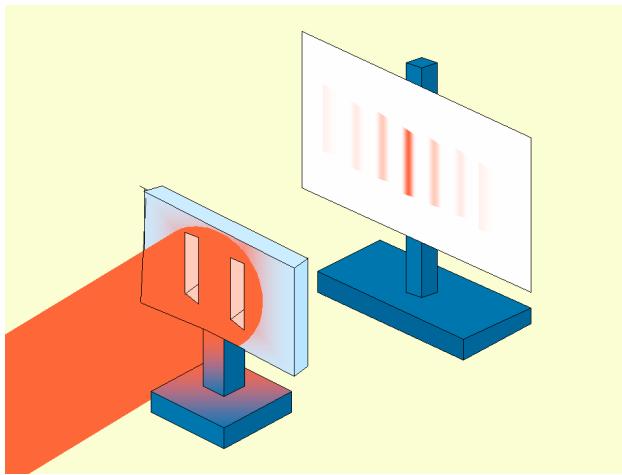


Figure 5.1: A sketch of the classical double-slit experiment. A coherent source of light is shone on a barrier containing two slits the light can pass through. The result is an interference pattern in the detection plane. A photo of the resulting interference pattern (as well as the single-slit case) is shown in Figure 5.2. This image is taken from March [53].

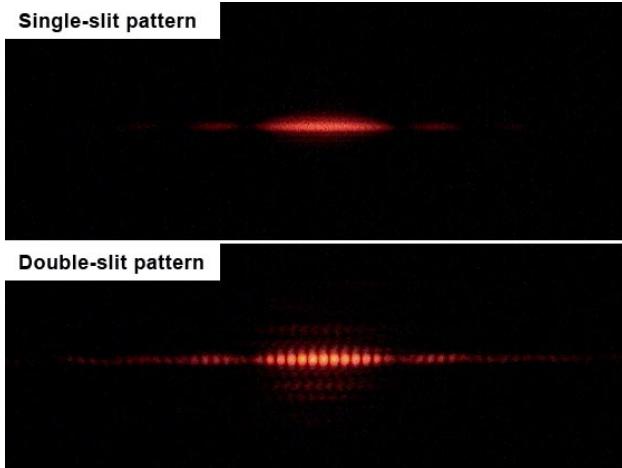


Figure 5.2: Photos of single-slit diffraction and double-slit interference patterns generated by laser light. The single-slit results in a sinc function pattern. The double-slit results in a sinusoidal pattern within a sinc envelope. The envelope is due to the non-zero slit widths. These photos are taken from Jordgette [54].

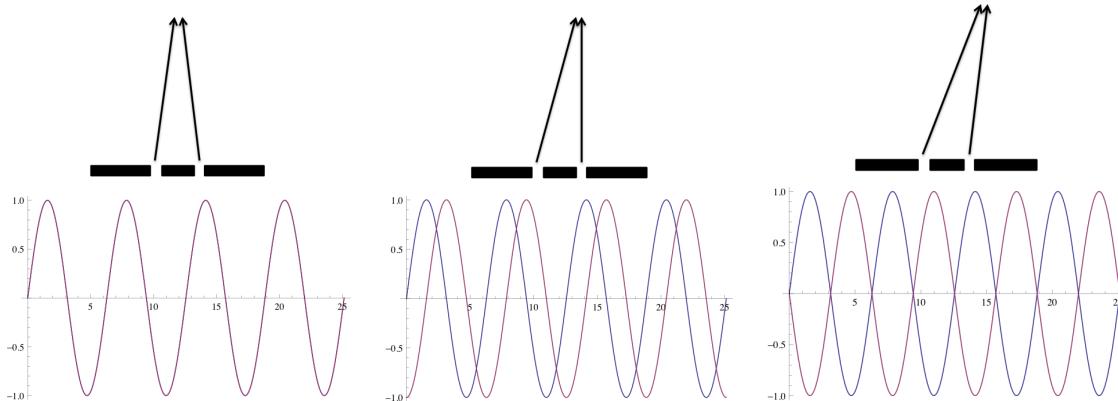
equations. This implies that they usually obey a superposition³ principle, i.e. that the value of the wave at a given point is equal to the sum of the values of all component waves at that point. When the waves from the source reach the double-slit, most of the wave-front is blocked and only two points are able to pass through. These two points emanate waves outward in all directions. Due to the superposition principle and differences in path lengths between waves from each slit, an interference pattern forms in the total wave. It should be clear from the figure that the details of this pattern must depend upon the distance to the detection plane, the wavelength of the wave, and the separation between the slits. Think about these things for a while, and you'll realize that the width and spacing of the bands increases with: (1) larger distance to the detection plane, (2) larger wavelength, and (3) smaller spacing between the slit. Likewise, the opposite is true for the opposite variations.

This effect can be demonstrated for any waves that obey a superposition principle. For example, as shown in Figure 5.4, the effect can be seen in water ripples. (A single-slit diffraction pattern envelope is often seen due to the non-zero width of each slit. Although they are not of great concern to this paper, the details of this pattern can be explained by a generalization of the double-slit argument where each single slit is thought of as infinitely many sources side-by-side.⁴)

This experiment was at the heart of the wave theory of light. The wave theory of light went through some subtle changes regarding the question of what medium (if any) light waves were transmitted through, before settling on the picture attributed to Maxwell in which light is a wave disturbance of the electric and magnetic fields. Little did contemporary physicists know, the double-slit experiment would find a new life as an important phenomenon in the description of “particle” dynamics.

³Sound familiar?

⁴An interesting question is what effect does the *height* of the slits have. This can also be determined using a generalized argument and, again, is not of concern to this paper.



(a) The path lengths are equal, causing the two waves to completely overlap and interfere entirely constructively.

(b) The left path length is a bit longer than the right path length, causing the left wave to lag behind the right wave. Partial destructive interference occurs.

(c) The left path length is half a wavelength longer than the right path length, causing the left wave to lag behind the right wave by exactly one-half of a period. Total destructive interference occurs.

Figure 5.3: A rough graphical argument showing how a wave theory is sufficient to describe a double-slit interference pattern. Each subfigure analyzes a particular point in the detection plane with respect to the waves emanating from the two slits. There are diagrams for a light band, a medium band, and a dark band (left to right). Each double-slit diagram depicts a bird's-eye view of the apparatus. From this perspective, the source (omitted) is below the diagram and the detection of the pattern takes place towards the top of the diagram. Below each double-slit diagram is a graph corresponding to the value of the wave at the detection point over time. The purple plot corresponds to the left wave and the blue plot corresponds to the right wave. The vertical axis of each graph represents the wave value. (One might think of this as a component of the electromagnetic field, for instance.) The horizontal axis of each graph represents time. The starting time $t = 0$ is not necessarily the same for all three graphs. The total wave value at the detection point is understood to be the sum of the values of the waves emanating from each slit. Note: the terms leading and lagging can be confusing when looking at a time graph. In the center diagram, the purple wave *lags* the blue wave because the first peak of the blue wave arrives at a time $t_{\text{blue}} \approx 1.5$ while the corresponding peak of the purple wave doesn't arrive until the *later* time $t_{\text{purple}} \approx 3$.

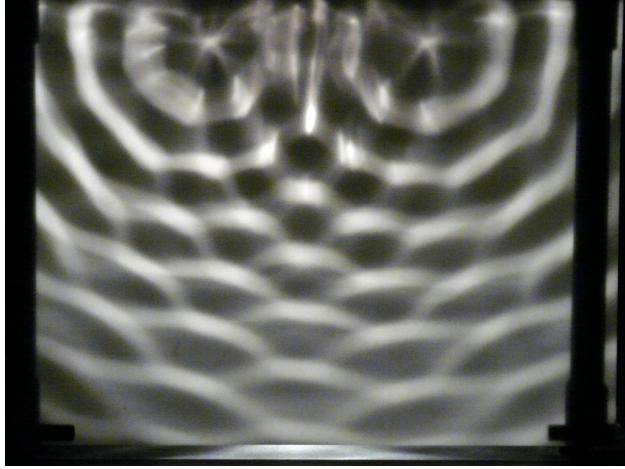


Figure 5.4: A demonstration of double-slit type interference of water ripples. Two point sources are created by periodically disturbing the water at two points. The waves in the water propagate outwards forming an interference pattern equivalent to a double-slit interference pattern. An image of the water ripples is projected onto a screen. This is what is shown in the figure. The photo is taken from MIT [55], where a video of the process is also available. The same effect can be demonstrated by blocking the waves from a single disturbance with a proper double slit.

5.2 The Quantum Double-Slit Experiment

In the early twentieth century, a number of issues including blackbody radiation and the photoelectric effect lead physicists such as Planck [56] and Einstein [57] to propose that light had a particle-like nature. For various reasons, these phenomena were unintelligible in terms of the classical wave theory of light.⁵ These particles of light came to be known as “photons”⁶, and were understood to be necessarily massless due to relativity. Thus began

⁵Although Einstein [57] provided the best-known explanation of the photoelectric effect using particles of light, it was later shown by Lamb and Scully [58] that one need not appeal to such quantization of light. Instead, they showed that it sufficed to introduce a quantum theory of matter. Thus, the photoelectric effect *on its own* does not preclude a classical wave theory of light. Many other phenomena, however, *do* require this quantization in order to be explained.

⁶The term photon was coined by Lewis [59].

the age of treating light as a particle whose dynamics are dictated by quantum mechanics.⁷

This seemed at odds with the results of Young's double-slit experiment which suggested that light was a wave. How could *particulate* light form an interference pattern? If we shoot *classical* particles at a double slit, they would eventually form two blobs on the other side, one centered about each slit. The two blobs would combine to form one big blob. That is, classical particles do *not* form an interference pattern. This is the whole reason why the interference pattern formed in Young's double-slit experiment convinced people in the nineteenth century that light was a wave. However, this apparent contradiction in the nature of light needed to be reconciled. A host of speculations arose: Maybe light is only wave-like when a ton of photons are being shot at the double slit? Maybe we would see the classically predicted big blob of particles on the other side *only* if we shot single photons at a double slit *one at a time*.

Such single photon double-slit experiments were performed by Taylor [61] and others. The results were shocking. One particle at a time was detected, and each particle landed

⁷This is perhaps the most misleading sentence in this paper. It wasn't until later that formal quantum theories of light came to fruition in the form of quantum electrodynamics and, later, other quantum field theories. The modern understanding of light is that it is a disturbance in a quantized field. One begins with the classical electromagnetic four-potential A^μ and promotes it to a quantum operator (more accurately a set of quantum operators, one for each point in space). These operators are analogous to observables that represent "measurements" of the field at that point. The set of distinct classical field configurations become orthogonal quantum states. The dynamics of light alone can be then given in the path-integral approach using the classical Maxwell Lagrangian $\mathcal{L} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu}$. Here the integral over all paths becomes an integral over all possible classical histories for the field, with fixed starting and ending field configurations. (One could think of this as doing a separate path integral for the field value at each position.) That's an uncountably infinite number of independent traditional path integrals. The details of the quantization force the energy of the field to take discrete values, and it is these "field quanta" that get interpreted as photons. More generally, quantum field theories replace every elementary particle by some field which is then quantized according to quantum theory, and field quanta become interpreted as the "particles". The Lagrangians of quantum field theories combine the Lagrangians for the individual particle fields and then add terms for interactions between certain groups of these particle fields (for example, the interaction of the electron and photon fields). For reasons that aren't entirely clear to me, we still get away with thinking of single independent particles that have some individual quantum states for a surprisingly large number of circumstances. Even though there are all sorts of problems with treating photons as particles with wavefunctions (see Newton and Wigner [60]), one can often get away with such a loose description in the context of almost any lab, shy of those with particle accelerators. In this paper, I don't concern myself with such details outside of this footnote, and I will treat light as made up of some particles called photons that have some sort of wavefunction. The indistinguishability of these particles is a whole other can of worms.

at some well-defined position in the detection plane. Over time the traces of the particles built up, only to exhibit the same old double-slit interference pattern! Physicists then turned their attention to objects that they felt were more certainly particulate, such as electrons. What happens if we shoot electrons at a double-slit one at a time? Jönsson [62] performed such an experiment in 1961, and the result was the same buildup of a double-slit interference pattern! This result was expanded upon by Tonomura et al. [63] in 1989 by ensuring that there was no more than one electron in the apparatus at any given time. The buildup of the interference pattern they found is shown in Figure 5.5.

This same procedure of “Well what about *these* particles?” was rinsed and repeated for increasing larger particles such as atoms and even molecules. These experiments continued to demonstrate the signature fringe pattern of interference. A notable example was a double-slit experiment using buckyballs⁸ by a team led by Zeilinger, see Arndt et al. [64], in 1999. Another team which Arndt co-led set the current record in an article published this past March, Juffmann, et al. [65]. They successfully demonstrated double-slit interference using molecules of $C_{48}H_{26}F_{24}N_8O_8$, a molecule of mass 1298 amu containing 114 atoms.⁹

So fine, we are forced to acknowledge that even relatively large (composite) particles exhibit double-slit interference. But how does it work? Quantum mechanics, of course. A very hand-wavy argument in terms of the path-integral formulation is as follows: At the double slit, each particle is in a superposition of passing through the left and right slits. The probability of detecting a given particle at some point in the detection plane is given by the magnitude squared of the sum of the amplitudes for every possible path it could take to get there. The amplitude for a path is given by taking the action of the path divided by \hbar to be the phase, and taking the magnitudes of the amplitudes for all paths to be the same (see Equation 1.27). The actions for these paths are much larger than \hbar . Therefore, only paths

⁸A cute term for buckminsterfullerene C_{60} , which is essentially a soccer ball skeleton of carbon atoms.

⁹Say what!? That’s right. 1298 amu, 114 atoms. For real. The pattern buildup was even recorded in real-time video!

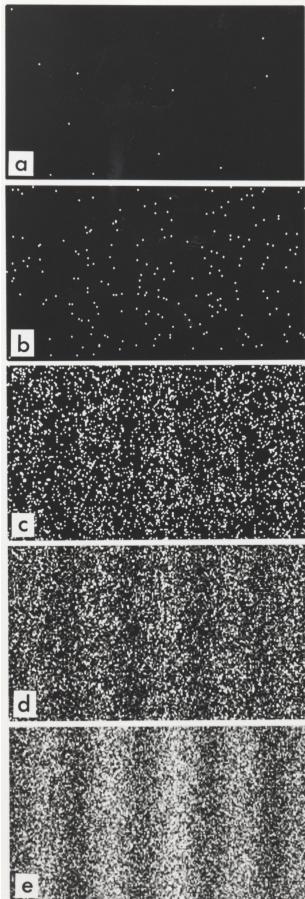


Figure 5.5: The buildup of a double-slit interference pattern from single electrons. In this experiment performed by Tonomura et al. [63], single electrons are fired one at a time at a double slit. The buildup of a double-slit interference pattern illustrates the wave-like properties inherited by particles due to quantum mechanics.

close to the classical paths of stationary action contribute significantly to the probability. For any other path, nearby paths will vary greatly in phase, with the result that all other paths roughly cancel each other out. The stationary action paths are precisely the straight line paths that connect either slit to the detection point in question. Longer paths have larger actions, causing a periodic change in the phase associated with each path. The situation is now analogous to the classical picture from Figure 5.3, but with a small change: instead of adding oscillating real numbers, we add complex numbers of modulus 1 that rotate around and around in the complex plane. The argument is essentially the same: differences in the path length cause relative phase differences that lead to constructive interference in some cases and destructive interference in others. The main difference is that now the interference lies in the probability distribution of finding the particle at different points instead of some wave value. (A more formal treatment of the situation in terms of momentum basis analysis is given by Marcella [66].)

Particles only build up in a double-slit interference pattern if they are in a superposition of traveling down the left and right slits. A pattern will not form from particles in a mixture of traveling down the two individual slits. For the traditional pattern, the particles must all be in some *equal* superposition of traveling down the left and right slits $\frac{1}{\sqrt{2}}(|L\rangle + e^{i\theta}|R\rangle)$ which is allowed to have some relative phase shift¹⁰ θ between the two arms. Different relative phase shifts produce patterns whose bands¹¹ are shifted left or right by some amount.

The state $|+\rangle = \frac{1}{\sqrt{2}}(|L\rangle + |R\rangle)$ produces a pattern with a bright band in the center of the detection plane. The state $|-\rangle = \frac{1}{\sqrt{2}}(|L\rangle - |R\rangle)$ produces a pattern that is shifted so that the bright bands correspond to the dark bands of the $|+\rangle$ pattern, and vice-versa. Any two such states that differ by a π phase shift produce patterns that are opposite to each other. This highlights an interesting fact. Consider the state of a particle in an equal mixture of

¹⁰It has to be the *same* relative phase shift for every particle in order for an interference pattern to build up.

¹¹I mean the underlying double-slit bands, not the single-slit envelope.

traveling down the left and right slits

$$\rho_{\text{mixture}} = \frac{1}{2} |L\rangle\langle L| + \frac{1}{2} |R\rangle\langle R|. \quad (5.1)$$

If you replace $|+\rangle$ and $|-\rangle$ by the appropriate linear combinations of $|L\rangle$ and $|R\rangle$, you'll see that this density operator can be rewritten as

$$\rho_{\text{mixture}} = \frac{1}{2} |+\rangle\langle +| + \frac{1}{2} |-\rangle\langle -|. \quad (5.2)$$

Thus, this state is entirely equivalent to the particle being in an equal mixture of the $|+\rangle$ and $|-\rangle$ states.

Which is it then, does the particle interfere or not? If you think of the particle as being in a mixture of $|L\rangle$ and $|R\rangle$, then the two blobs from each slit add up and we get one big blob at the detection plane. Alternatively, if you think about it as a mixture of $|+\rangle$ and $|-\rangle$, the two interference patterns are just plopped on top of each other and you get the *same* big blob in the detection plane. (Well of course you do: it's in the same physical state!) The point is that it isn't really meaningful to ask "Which mixture is the particle 'really' in?" They all correspond to the same physical state, and they all produce the same big blob. This fact will reappear in Part II.

More generally, this density operator is the identity operator divided by two

$$\rho_{\text{mixture}} = \frac{I}{2}. \quad (5.3)$$

We know from Equation 1.12 that the identity operator can be decomposed into an equal mixture of *any* set of states that form an orthonormal basis for the Hilbert space in question. In the double-slit context, this implies that we can think of ρ_{mixed} as an equal mixture of any two orthogonal states. Such states will always be guaranteed to generate complementary

patterns that add up to the one big blob.

All this talk about the double-slit experiment for general particles might have you wondering: “Why then don’t I see baseballs form an interference pattern after throwing them through two open windows?” Well, for one, it’d be really hard—probably impossible—to get baseballs in coherent superposition states. As I discuss in the next section, any sort of information that can lead to discerning particle-like behavior necessarily screws up the required coherence. Second, even if we *could* put baseballs in superposition states, the details of quantum mechanics dictate that the interference pattern they would create would be made of light and dark bands so closely packed together that we could never expect to be able to distinguish them from each other with any realistic measurement.

5.3 Complementarity

The sort of hocus-pocus argument that some physicists gave (and still often give) to explain the quantum double-slit experiment is that each particle travels down both slits simultaneously.¹² This bothered people that had classical biases about the way particles ought to behave. Such people believed that it was possible, in principle, to show that each particle goes through one slit or the other, but not both. So they constructed experiments where they would tag photons depending on which slit they went through. The intention was that after they had collected the interference pattern, they could point to the tags on each photon and show which individual slit each photon came from.¹³ The surprise for them was that an interference pattern no longer built up whenever they tagged photons at each slit, and it didn’t matter how they tried to tag them.

To see why this is the case, consider the following simplified example: A particle begins in the state $|+\rangle$ from the last section. If we fire such particles at the double-slit, we see

¹²I discussed why this is nonsense on page 15.

¹³I’m greatly exaggerating the details, but I promise you the heart is the same.

interference. Next, we tag each particle depending upon which slit it goes through. We could envision this process as tensoring in a subsystem corresponding to the tag on the particle. After each $|+\rangle$ particle passes through the taggers, their state becomes

$$|\psi\rangle = \frac{1}{\sqrt{2}}(|L\rangle \otimes |\text{This one's a lefty!}\rangle + |R\rangle \otimes |\text{This one's a righty!}\rangle). \quad (5.4)$$

If our tags are any good, then they won't have any effect on whatever underlying mechanism causes the interference.¹⁴ Thus, the effective double-slit state for the particle is given by $\text{tr}_{\text{tag}}(|\psi\rangle\langle\psi|)$. Now go ahead and take this partial trace. What you'll find is that the reduced density operator is an equal mixture of the $|L\rangle$ and $|R\rangle$ states, i.e. the identity operator divided by two. We know what pattern this state produces; the big blob, no interference! However, if we used the same tag for both slits (a tag that is useless for figuring out which slit the photon went through), we find the marginal state to be just $|+\rangle$ by the linearity of the tensor product, and interference returns. For tag values that can be superpositions of the completely distinguishable values (such as the semi-quantum treatment of polarization), we can come up with an intermediate case in which we have tags that are partially useful in determining which slit each photon went through. We then end up with a pattern that is somewhere between the perfect interference pattern and the big blob.

This suggested to physicists that we can either have so-called “which-slit information” or we can have so-called “interference information”: the more information we have for one, the less we have for the other. This is a manifestation of Bohr’s complementarity¹⁵ principle,

¹⁴In reality, polarization tended to be used as a tag. From a deeper, field theoretical framework of the photon, polarization has *everything* to do with the interference mechanism, and the tags directly screw up the pattern. This is true classically too. So, polarization is a pretty lousy choice. The argument I make for why tagged particles don’t interfere, however, roughly shows why there is no better tag that could suffice. In the semi-quantum view of the photon that we often take, you could treat polarization as some tag the photon is born with. Tagging a photon then corresponds to merely modifying its polarization, and the argument as to why they don’t interfere remains the same.

¹⁵Not to be confused with Bohr’s *correspondence* principle, which involves how we recover classical physics from quantum physics in the everyday limit. I used to always mix up the two names.

which roughly states that the results we get from measurements intrinsically depend upon the type of measuring device, and that there are certain complementary natures of objects such that an experiment can reveal one or the other, but not both at the same time with perfect clarity. In the double-slit context, we can setup our apparatus to reveal the particle-like nature of photons (which-slit), or we can set it up to reveal the wave-like nature (interference), but we cannot observe both natures simultaneously with perfect clarity.

One should be careful when dealing with such interpretive concepts, as they often mix intrinsically classical notions with what is an entirely quantum situation. However, the complementarity of the double-slit can be derived in a formal context where it is known as the Englert–Greenberger duality relation¹⁶

$$D^2 + V^2 \leq 1 \tag{5.5}$$

where D is a formal measure of the distinguishability of which slit the photon went through and V is a formal measure of the interference visibility. This relation was first demonstrated experimentally by Greenberger and Yasin [67], later proven theoretically by Jaeger, Shimony, and Vaidman [68], and independently by Englert [69]. Englert was explicit about the fact that his derivation was completely independent of the uncertainty principle. Although I am not familiar with the formal details of the Englert–Greenberger relation, I would argue that

¹⁶No experiment can violate the proper complementarity principle. I am not an expert on this relation, so I can't endorse it as such, but I strongly believe it is the proper formal version. Do not take any experiment that claims to violate this at face value; be very careful in your scrutiny. In particular, there is one experiment out there that claims to violate the relation and subsequently complementarity using an ingenious setup involving a wire grid. However, the experiment is severely flawed on a number of counts. I would love to explain it here, as it is truly a clever setup and an excellent demonstration of how well one can hide flaws and how stringent scientific scrutiny must be. But having said that, I unfortunately must refuse to even mention the name of this experiment. The author, who has very little qualifications and presents them misleadingly, has been very stubborn and closed minded when the scientific community explained the experiment's flaws to him. In turn, he has made outrageous accusations. I leave it to the interested reader to find the experiment on their own and to discover the flaws. Again, it involves a wire gird (and a double slit). There is ample criticism out there that explains the flaw properly, but beware: there is also some criticism by reputable physicists that incorrectly identifies flaws that don't essentially apply.

it is probably a manifestation of the uncertainty principle (at least in spirit) because the measurements of which-slit and interference information are closely related to observables that do not commute. That is, the states that generate perfect interference are different from those that have well-defined which-slit information. As a result, these measurements ought to obey a Robertson uncertainty relation that puts bounds on their statistics for any given state.

In this context, an interesting question comes up: What happens if we gain which-slit information (say by tagging the photons), but then erase it? Can we regain an interference pattern? These questions are the motivation of so-called “quantum eraser” experiments. Such an experiment is the subject of Part II, and I discuss other historical examples as well.

Part II

The Experiment

Preface to Part II

This part gives an account of the title experiment. The reader is not assumed to have explicitly read Part I, but familiarity with the concepts introduced there is necessary. This part can be taken to stand entirely on its own. However, it does not take the form of a standard journal article describing such an experiment. It is written at a level understandable to someone with a minimal understanding of quantum theory and is accordingly not deeply technical. It makes many simplifications in order to reach this audience.

Chapter 6

Inspiration

In this chapter, I diverge from the otherwise more formal nature of Part II to give a personal account of how I arrived at my experiment. I discuss other experiments I have worked on in my undergraduate career, and the thought process that led to the proposal of the experiment discussed in this paper.

6.1 Old Ideas

In the first few days of my undergraduate career, I asked my advisor how I could begin learning quantum mechanics. He pointed me to the approachable book by Greenstein and Zajonc [70]. After reading it, I proposed a double-slit experiment where one would use two lasers, one aimed at each slit with beams narrow enough that they didn't overlap the other slit. I then wondered what would happen if only one photon were in the apparatus at any time. Clearly, the photon would have to come from one of the two lasers and thus could only have gone down a single slit. If one could generate an interference pattern from this experiment, it would clearly violate Bohr's complementarity, or so I thought.

In reality, you'll only see an interference pattern if the photons are completely indis-

tinguishable without any sort of tag that separates them.¹ When you have this condition, then each single photon is effectively in a superposition of coming from each source. The experiment wouldn't prove anything more than the fact that quantum physics is the physics governing the generation of the photons too, which is obvious. Moreover, Pfleegor and Mandel [71] had already done this experiment in 1967. (It was even in Greenstein and Zajonc [70]!) The only difference was that they didn't use a double-slit. They generated interference by crossing the beams at a detection plane. Somehow, I was convinced that doing it with a double-slit was different. It's not really.

I could have proposed to do it with better single photon sources than the simply attenuated lasers they used, but this too had essentially already been done by de Riedmatten et al. [72] and Kaltenbaek et al. [73]. Nevertheless, I carried on with my attenuated laser version, and I never really got anywhere with it. (It's not exactly an easy setup to align, and that's probably why no one ever bothered to do it exactly this way.) It's not a totally useless proposal, though. It would still be interesting to see it done, if for no other reason than the historical importance of the exact double-slit context. Some young student who eventually takes my place in my advisor's lab will probably pick up where I left off. They just have to keep in mind that it isn't as big a deal as I once thought it was.

I came up with some other ideas in my early days too. Most of them had to do with trying to violate complementarity using entangled photons. In any event, once I studied quantum theory from a mathematical standpoint and learned that entangled states always had mixed marginals, it was clear why these experiments wouldn't work and why you could never use entanglement to violate complementarity. This is the pitfall of relying on a non-mathematical background in quantum theory. You come up with all sorts of ideas that are either obvious or silly from a mathematical standpoint. That's because quantum mechanics is so counterintuitive. You can't really build up a good intuition for how it works until you

¹See my discussion in Section 5.3.

look at the math.

6.2 A New Idea

At the same time that I realized why my last idea was silly from a formal perspective, I came up with a new idea sometime at the beginning of my sophomore year, which is the idea that forms the basis of this paper. It was inspired by what is essentially known as “steering”, suggested by Schrödinger [74] and later formalized by Wiseman, Jones, and Doherty [75, 76].

Consider the entangled state of two photons² given by

$$|\Phi^+\rangle = \frac{1}{\sqrt{2}}(|L_1\rangle \otimes |L_2\rangle + |R_1\rangle \otimes |R_2\rangle) \quad (6.1)$$

where $|L_1\rangle$ and $|L_2\rangle$ are the states of a photon passing down the left or right slit of some double-slit, and $|R_1\rangle$ and $|R_2\rangle$ are the states of a photon passing down the left or right slit of some other double-slit. As you can see, if the first photon passes down the left slit of its double-slit, the other will pass down the left slit of its own double-slit. Likewise for the right slit case. Let’s consider the results of the experiment being carried out on the first photon. If we calculate the reduced density matrix, we find $\text{tr}_B(|\Phi^+\rangle\langle\Phi^+|) = \frac{I}{2}$. This photon is effectively in a mixture of going down either slit, so we won’t see an interference pattern. (In the context of complementarity, one might say we won’t see interference because the other photon contains which-slit information about the first photon due to their entanglement.)

Equivalently, the first photon is effectively in a mixture of the equal superposition $|+_1\rangle = \frac{1}{\sqrt{2}}(|L_1\rangle + |R_1\rangle)$ and the phase-shifted equal superposition $|-_1\rangle = \frac{1}{\sqrt{2}}(|L_1\rangle - |R_1\rangle)$. We could then think of the photon as being in a mixture of interfering one way and interfering in the opposite way so that the two patterns add up to the same big non-interference blob. A quick

²The details of Bose–Einstein statistics require that we symmetrize such states, but this would just complicate the point. A better analysis would use Fock states, but this is beyond the scope of this paper.

change of basis on $|\Phi^+\rangle$ shows that you can rewrite the bipartite state as

$$|\Phi^+\rangle = \frac{1}{\sqrt{2}}(|+1\rangle \otimes |+2\rangle + |-1\rangle \otimes |-2\rangle). \quad (6.2)$$

We see that if we make a measurement that finds the second photon in the state $|+2\rangle$, then the first photon will also be in the equal superposition state and will interfere. Likewise, if we make a measurement that finds the second photon in the state $|-2\rangle$, then the first photon will also be in the phase-shifted equal superposition state and will interfere in the opposite way.

In other words, by measuring the second photon in the $|+2\rangle, |-2\rangle$ basis we can “steer” the first photon into the corresponding basis. Of course, such a measurement would return either state with a 50% chance. Half the time we would see one interference pattern; the other half of the time we’d see the opposite interference pattern. Thus in total, we’d still just see a big blob. The difference, however, is that such a measurement would tell us which photons should form which pattern. We could then use this information to regain the interference pattern by keeping only those detections in the first arm that occur when the second photon is found in the equal superposition state.

The question that remains is this: How do we make such a measurement of the second photon? My idea was simply to use a Hermitian 50-50 beamsplitter (essentially a half-transparent mirror). If we take $|L_2\rangle$ and $|R_2\rangle$ to be the states corresponding to the a photon going down either the left or right input arms of the beamsplitter, then the unitary gate imposed on the state by the beamsplitter is given in the $|L_2\rangle, |R_2\rangle$ by

$$U = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \quad (6.3)$$

where, after the gate is applied, $|L_2\rangle$ and $|R_2\rangle$ are understood to be the states corresponding

to a photon being in either the left or right output arm. What we see is that $U|+_2\rangle = |L_2\rangle$ and $U|+_2\rangle = |R_2\rangle$. We can then replace the double slit in the second arm with a mirror edge that splits the beam into the parts that used to correspond to the left and right slits. We can then use mirrors to feed these parts into the beamsplitter. This would transform the superposition states into left path or right path states. The effect on the two-photon state is given by

$$(I \otimes U)|\Phi^+\rangle = \frac{1}{\sqrt{2}}(|+_1\rangle \otimes |L_2\rangle + |-_1\rangle \otimes |R_2\rangle). \quad (6.4)$$

We can then simply count the first-arm photons only when we detect a photon in the left output port of the beamsplitter in the second-photon arm. This corresponds to a measurement result that collapses³ the composite state onto $|+_1\rangle \otimes |L_2\rangle$, and thus the first-photon marginal state onto $|+_1\rangle$, which would generate an interference pattern.

This is an example of what is known as a “quantum eraser” experiment because we “erased” the which-slit information that the second photon had about the first photon. This erasing was done by the beamsplitter, after which no measurements of the second photon could discern any which-slit information about the first photon. (I’ll discuss the extent to which this has been done before in Chapter 7.)

All of this can be, in principle, carried out in the lab. However, the problem is that it’s just too obvious. Even someone who has just learned quantum mechanics could see what would happen and why. (Heck, a *sophomore* figured it out!) Hopefully, even some readers of this very paper who began with no experience in quantum mechanics are able to understand what would happen and why. Decades ago, this experiment would have been groundbreaking. (1) It would have been very counterintuitive because quantum mechanics wasn’t as well understood, and (2) it would have been a significant test of quantum mechanics. At this point however, quantum theory has been proven to be valid to within any reasonable boundaries

³Traditional photon detections are excellent examples of where the von Neumann measurement formalism (shown in Sections 1.2 and 3.2) fails. A more technical treatment would use the POVM formalism, but the difference is inconsequential for the broader point.

we could define for this situation. Furthermore, the math is now so well understood that this result is somewhat obvious.

But that's the the virtue of being an undergrad. You're not expected to produce some groundbreaking piece of work. It's enough to just play with these things until you get a feel for how they work. And even when you think you *know* how an experiment will end up—that is, when it does seem obvious—there is still value in getting in the lab and actually physically doing it first hand.

I am a bit embarrassed now, looking back at some of the things I proposed and some of the things I thought would be significant. But I'm not regretful, by any means; instead, it seems to me like a good measure of success. When you look back at the you from only a few short years ago, and the person you see makes you want to palm the face you currently have, it just shows you how much that person has managed to learn and improve in such a short amount of time.

Chapter 7

Previous Experiments

In this chapter, I give a brief review of previous experiments similar to the one this paper is based on. I explain how my experiment differs from those that have already been performed.

7.1 Quantum Erasers

The proposal for a quantum-eraser experiment—such that one obtains which-slit information about particles, but then subsequently erases it and regains interference—is due mostly to Scully [77, 78, 79, 80]. For a while, experimental details prevented this from being carried out. A number of variations were proposed that had their own difficulties associated with them. Finally, in 1992, the first successful experimental quantum eraser was performed by Kwiat, Steinberg, and Chiao [81].

This experiment did not use the double-slit context. The first double-slit quantum-eraser experiment was performed by Walborn et al. [82] in 2002. Their apparatus sketch, along with an explanation of the experiment, is shown in Figure 7.1. In their experiment, they used circular polarization to tag photons passing through either slit of a double slit. These photons were polarization-entangled with other photons that were detected in another arm. It should

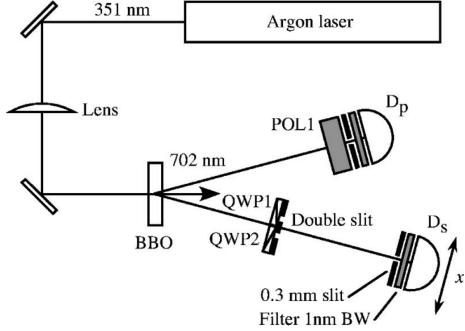


Figure 7.1: The double-slit eraser performed by Walborn et al. [82]. The picture is taken from their paper. Polarization-entangled photons are generated by using a β -barium borate (BBO) crystal. Quarter-wave plates (QWP) with perpendicular axes are used to alter the polarization of photons passing through the double slit depending upon which slit they pass through. This results in one of two circular polarizations, introducing which-slit information. In the other arm, a linear polarizer collapses the state of the other photons onto one that has no circular polarization information. The collapse induced by the linear polarizer erases the which-slit information about the double-slit photons because of the entanglement between the double-slit photons and the other photons prior to their collapse. Interference is then observed in coincidence.

be emphasized here that the photons were not position-entangled. The second photon would have no which-slit information about the double-slit photon were it not for the polarization tagging, and one would expect interference there if there were no quarter-wave plates.

Without modification, this would not produce interference because the second-arm photons contain which-slit information about the double-slit photons. However, by passing the second-arm photons through a linear polarization filter (polarizer) before detection, which-slit information was erased, and the first arm produced interference. The key to this result is that each entangled pair was detected in coincidence. Thus, the only photons detected at the double-slit were those whose entangled partners had collapsed onto a specific linear

polarization. All of the other pairs are thrown out. If the linear polarizer used in the second arm had been rotated 90° , the opposite double-slit pattern would have developed, which Walborn et al. demonstrated.

A proposal by Wheeler [83] suggested that one could *choose* whether or not to erase the which-slit information *after* the double-slit photon has passed through the double-slit.¹ In fact, in an entanglement-based quantum eraser you could go so far as to choose whether or not to erase the which-slit information *after the double-slit photon has already been detected!* This was shocking to many people. How could you decide whether or not the photon interferes after you had already detected it?

If you really understand what's going on, you'll see that this isn't so surprising after all. No retro-causality comes into play. You're merely using coincidence² detection to pick out *which* of those photons have interfered in a certain way. (Well, at least, that's the best way of explaining it to someone fixed in a classical outlook. Doing the quantum math makes what's going on much clearer.) Regardless, I shouldn't make this effect seem too mundane. It is certainly a shocking result in its own right and marks an important point in the history of quantum theory. You cannot explain this effect in any classical way.

Walborn et al. [82] were aware of Wheeler's proposal, and they noted that their result was an example of delayed *erasure*. That is, they were able to put off detecting the second-photon until after the double-slit photon had already been detected. They still saw interference in coincidence, of course, but they did not delay the *choice* of whether or not to erase which-slit information. The linear polarizer had been put in before any photons were even generated. Others went on to refine this result and perform a true delayed-choice quantum eraser. In the next section I will discuss one such example.

¹The detailed experimental proposal is due to Scully and Drühl [77].

²This key detail is often forgotten by people, and it leads them to believe that quantum erasers can send information faster than light. This is not the case. You will not see an interference pattern at the double-slit unless detection information about the second photon is sent over to decide which double-slit photons to keep.

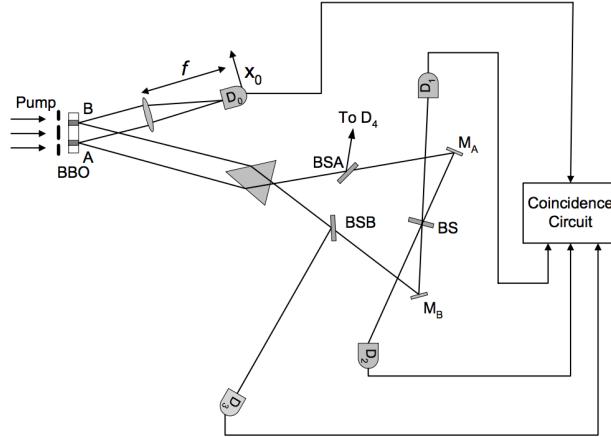


Figure 7.2: A sketch of the delayed-choice quantum eraser performed by Kim et al. [85]. A coherent pump beam is masked by a double slit and is subsequently incident on a BBO crystal. This generates a pair of photons that are entangled in their generation location. One photon is used to attempt detection of an interference pattern. The other is used to measure which-path information or erase this information by combining paths on a beamsplitter.

7.2 The Delayed-Choice Quantum Eraser

Delayed-choice quantum erasers had been performed as early as 1987, when Hellmuth et al. [84] published two different takes on the classic proposal. Their paper included both space and time domain experiments, but neither of these were in the double-slit context. A number of experiments have been performed since then, the most well-known of which is that by Kim et al. [85]. A sketch of their experiment is shown in Figure 7.2.

In their experiment, a coherent pump beam is incident on a type-II phase-matching β -barium borate (BBO) crystal. This process splits a single pump photon into two photons of twice the wavelength. A double-slit is used to mask the pump beam so that it is only incident upon two localized points on the crystal. This has the effect of generating a pair of photons that are entangled in their birthplace. Both will be generated from either of the two

locally pumped points on the crystal. One of these photons is used to attempt to generate an interference pattern. The other will necessarily have which-path information about the first one.

The which-path information can be erased by combining the two possible paths of the second photon on a beamsplitter and coincidence detecting the first photon with one of the output arms of the beamsplitter. Alternatively, the two paths of the second photon can be directly detected, leading to perfect which-path measurements of the first photon. Kim et al. implement this choice with two preliminary beamsplitters in the second photon’s possible paths. When hitting these beamsplitters, the photon will randomly either (1) split off from its path and be detected, leading to a which-path measurement; (2) or continue onwards to the combining beamsplitter, erasing which-path information. The paths of the second photon to each of its possible detectors is made much longer than the paths of the first photon. Thus, the second photon hasn’t decided whether or not to erase which-path information until *after* the first photon has been detected.

This isn’t a particularly convincing delayed-choice quantum eraser because the “choice” is made by a photon that has interacted with the photon whose nature is in question. However, other experiments have used more convincing choice mechanisms, such as the previously mentioned Hellmuth et al. [84]. Perhaps the best delayed-choice quantum eraser to date, Jacques et al. [86], made use of a choice made by a quantum random number generator that was relativistically separated from the first photon’s entry into the interferometer.

7.3 How This Experiment Differs

The experiment proposed by this paper is a quantum-eraser experiment that is explicitly set in the double-slit context. The primary difference between this experiment and the double-slit quantum-eraser experiment previously performed by Walborn et al. is that this

experiment makes use of path-type entanglement as opposed to polarization entanglement. Instead of “tagging” photons as they pass through a given slit, the path-type entanglement generated in this experiment directly entangles which-slit information: the photons in the eraser arm of the experiment are generated with intrinsic which-slit information about the double-slit photons, independent of any correlations in other properties such as polarization. Furthermore, a long-term goal of this work is to demonstrate successfully a delayed *choice* in the double-slit context, which Walborn et al. did not demonstrate. Thus, this experiment is more comparable to the delayed-choice experiment performed by Kim et al., which also made use of path-type entanglement.

Although a double-slit mask was used by Kim et al., their experiment was not quite a traditional double-slit quantum eraser. In their experiment, the double slit was only used to localize two photon-generation points. In this paper’s proposal, the double slit fulfills its traditional role of partially blocking a beam of photons. Additionally, this experiment uses the momentum entanglement generated by the spontaneous parametric down-conversion (SPDC) process, which is a direct consequence of the way SPDC conserves momentum. This is to be distinguished from the position-type entanglement used by Kim et al., which was merely a consequence of the spatial coherence of the pump beam across multiple points on the crystal.

Taken together, these facts significantly differentiate this experiment from the other quantum erasers that have been performed to date. The first goal of this experiment is to demonstrate that a quantum eraser can be performed using the apparatus described in Chapter 8. In Chapter 10, I discuss the future prospect of promoting this experiment to a delayed-choice quantum eraser.

Chapter 8

The Apparatus

In this chapter, I discuss the details of the experimental proposal. I begin by introducing a theoretical sketch of the apparatus that emphasizes the spirit of the experiment. I then discuss the physical realization of this apparatus and the practicalities of implementing it.

8.1 A Theoretical Sketch

A rough sketch of the proposed apparatus is shown in Figure 8.1. Using a BBO crystal pumped by a collimated source, momentum-entangled photons are generated via collinear type-II SPDC. (To give a proper treatment of SPDC would require more formalism than I have developed in this paper.¹⁾ As I mentioned in the last chapter, SPDC is a nonlinear process by which a single photon gets split into two photons of twice the wavelength, i.e. half the energy. Conservation of momentum dictates that the two momenta of the outgoing photons must equal the momentum of the incident photon. This results in correlations in the angular paths taken by the two outgoing photons. By pumping with a collimated

¹⁾ Interested readers should consult previously mentioned texts to become familiar with the quantum treatment of the harmonic oscillator. Then, one should be able to approach most papers involving SPDC (such as the previously cited quantum erasers), although one might first consider reading about SPDC in a quantum optics text. A truly formal treatment requires a quantum electrodynamical approach involving field operators.

source, each incident photon has approximately the same transverse momentum. Thus, the sums of the transverse momenta of each pair take the same value due to conservation of momentum. This is what essentially leads to the momentum entanglement. Depending on details involving the condition of the pump, the two-photon output state of the BBO can be made to approximate closely the EPR-type momentum-entangled state

$$|\text{EPR}\rangle = \int_{-\infty}^{\infty} |p, -p\rangle \, dp \quad (8.1)$$

where p is understood to be the transverse component(s) of the photons' momenta.

To make use of these momentum-entangled photons, a lens is used in a Fourier transform configuration. This configuration maps all of the different photon angles (momenta) onto different points on the double-slit, or equivalently, the mirror edge. A polarizing beamsplitter (PBS) is required to separate the entangled photons from each other. (Type-II SPDC generates photons with orthogonal polarizations, thus the PBS is the perfect tool to separate them.) After this point, the first photon moves on to the double slit, and the second on to the eraser. The eraser consists of a mirror edge and a 50-50 beamsplitter. The momentum entanglement and lens setup imply that the eraser-arm photon is reflected by the mirror edge when the double-slit photon passes through the left slit, and conversely, the eraser-arm photon misses the mirror edge when the double-slit photon passes through the right slit.² These two half-paths are then combined on the beamsplitter, and we detect those photons that leave the left output of the beamsplitter using a single-photon detector known as an avalanche photodiode (APD). The which-slit information is erased by this process of combining the half-paths and detecting a single output of the beamsplitter.

When this detector is triggered, it indicates that the state of the eraser-arm photon has collapsed onto an equal superposition state. We then detect the double-slit photon in

²Of course, this classical language isn't correct, but hopefully having read Chapter 6, it is clear to the reader what is really going on.

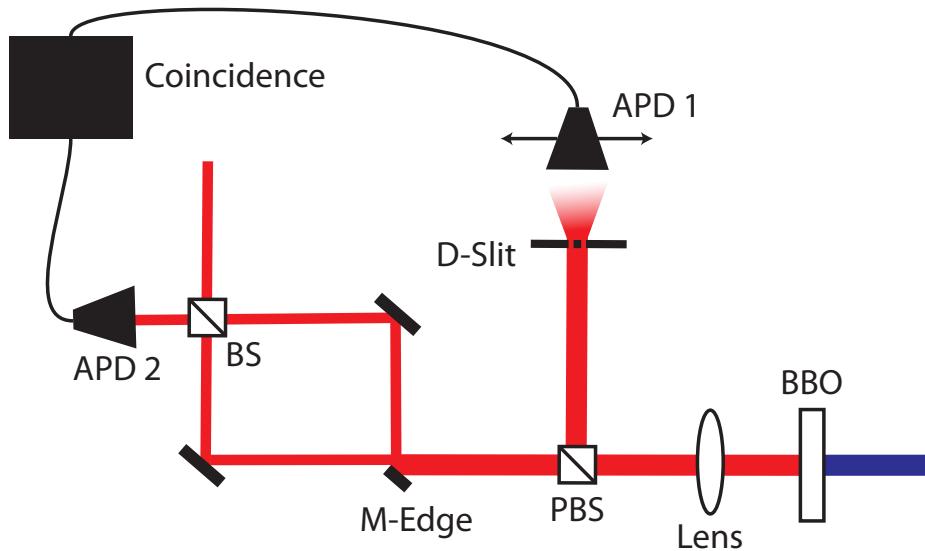


Figure 8.1: A sketch of the title experiment. Momentum-entangled photons are generated from a BBO crystal via SPDC. They are split apart by a polarizing beamsplitter (PBS). One photon is passed through a double slit, then detected with an avalanche photodiode (APD). The other photon has its profile split by a mirror edge (M-Edge). The two subpaths correspond to the two different slit paths of the first photon. The second photon paths are interfered on a beamsplitter (BS) and one output arm is detected. Coincidence detection is used to condition the double-slit photon pattern on detection of the eraser-arm photon. A lens is used so that the slit paths are determined by the photon's momentum.

coincidence, and expect to see an interference pattern because the entanglement dictates that the state of the double-slit photon must also collapse to a equal superposition state. Crucially, it is the same equal superposition for every double-slit photon, so we expect to slowly build up one pattern. Without having access to a single-photon detector array, we must scan the double-slit APD across the detection plane, building an image pixel by pixel.

SPDC is a very inefficient process. Provided our pump is weak enough, we can be quite confident that we do not have more than one pair of photons in the apparatus at any given time. This is in accordance with the spirit of the quantum double-slit experiment. It also introduces the practical issue of dealing with the photons that don't undergo the process.

8.2 Physical Realization

A photo of the physical realization of the proposed apparatus is shown in Figure 8.2. There are a number of items present in this realization that were not discussed in the simplified theoretical sketch. The apparatus begins with a 20 mW, 405 nm, continuous-wave laser diode module that functions as the pump laser. This is then collimated using a lens with a 100 mm focal length (Lens 1) followed by a second lens with a 50 mm focal length (Lens 2), before being incident on a type-II BBO crystal cut at 41.9° for degenerate SPDC pumped at this wavelength.

Next, the light passes through a prism to separate out the blue pump beam. All of the light exiting the BBO is collinear, and there is much blue light due to the inefficiency of the process. The prism splits the blue light off enough, so that it clips past the edge of a mirror and is mostly blocked. The remaining entangled photons then travel through a 400 nm far-field lens (Lens 3) in order for the positions at the double-slit and mirror-edge planes to correspond to momenta of the photons.

The entangled photons are finally split apart using a PBS and continue down separate

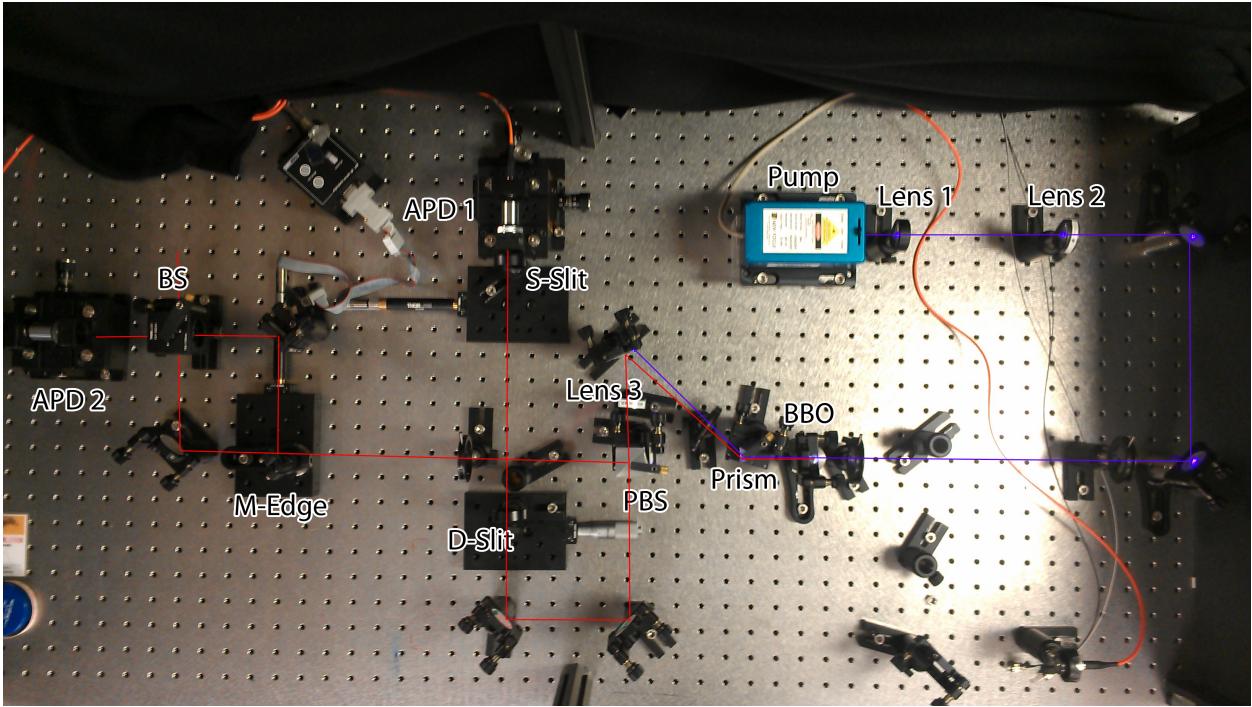


Figure 8.2: A photo of the physical realization of the title experiment apparatus. Colored lines are superimposed to represent the path of the pump laser and the path of the entangled pairs. A number of additions are made to the theoretical sketch in Figure 8.1 due to practical issues. These include the first lens, the prism, the single slit, and the filters (not labeled) placed before each detector. Coincidence detection is carried out using a computer that is not depicted. So too the APD's are themselves off-screen, but the microscope objectives used couple photons to the detectors are labeled in their place.

arms. The double-slit photons go on to pass through a double-slit, consisting of two 40 μm wide slits separated by a distance of 200 μm center-to-center. Instead of scanning an APD back and forth behind the double slit to detect a pattern one pixel at a time, we mount a single slit on a translation stage and scan *it* behind the double slit to pick out a local point on the pattern, which is then transmitted into a 10 \times microscope objective that focuses the light onto a fiber leading to the APD for detection. A band-pass filter, centered at 810 nm with a 10 nm bandwidth, is placed in front of the microscope objective to prevent extraneous photons from being detected. An empty mount immediately follows the double-slit to allow for the placement of a fourth lens with a 200 mm focal length, which may make the pattern easier to detect in the detection plane.

In the other arm, the photon is incident on the edge of a mirror correlated to the slit positions of the double-slit. This splits the path into two subpaths that are each reflected by other mirrors and then recombined on a symmetric beamsplitter. Finally, one of the output arms of the beamsplitter is also passed through a band-pass filter and a 10 \times microscope objective before being detected by another APD. The two APD's are managed via a connector block, timer board, and National Instruments™ LabVIEW™ software on a PC. This takes care of the necessary coincidence detection. The detection time window of each APD is 25 ns.

The ideal analysis presented in Chapter 6 made use of a Hermitian beamsplitter, but real experiments such this one often use symmetric beamsplitters. So, I must now revisit the analysis. The transformation of such a beamsplitter is given in the left-path, right-path basis³ by

³I should emphasize here that the “left” output arm of the beamsplitter is the arm reached by starting in the *left* input arm and being *reflected* by the beamsplitter or by starting in the *right* input arm and being *transmitted*. Conversely, the “right” output arm of the beamsplitter is the arm reached by starting in the *left* input arm and being *transmitted* by the beamsplitter or by starting in the *right* input arm and being *reflected*.

$$U = \frac{1}{\sqrt{2}} \begin{pmatrix} i & 1 \\ 1 & i \end{pmatrix}. \quad (8.2)$$

At first, it might also seem that the slit-correlated state used in the previous analysis (see Equation 6.1) isn't valid in light of the momentum anticorrelation of Equation 8.1, and that it should be replaced by the slit-anticorrelated state

$$|\Psi^+\rangle = \frac{1}{\sqrt{2}}(|L_1\rangle \otimes |R_2\rangle + |R_1\rangle \otimes |L_2\rangle). \quad (8.3)$$

However, every time a beam hits a mirror, the right and left sides get swapped. Looking at Figure 8.2, it is clear that the double-slit arm beam is reflected twice⁴, and thus has the same orientation it began with. The eraser arm, however, is only reflected once (by the PBS) before it reaches the mirror edge. Thus, left and right are swapped. For this reason, the correlated state from my old analysis is still valid. A similar analysis then shows that

$$(I \otimes U)|\Phi^+\rangle = \frac{1}{\sqrt{2}}(|L_1\rangle \otimes |-i_2\rangle + |R_1\rangle \otimes |+i_2\rangle) \quad (8.4)$$

$$= \frac{1}{\sqrt{2}}(|-i_1\rangle \otimes |L_2\rangle + |+i_1\rangle \otimes |R_2\rangle) \quad (8.5)$$

where the second line follows from some algebra, $|+i\rangle = \frac{1}{\sqrt{2}}(|L\rangle + i|R\rangle)$, and $|-i\rangle = \frac{1}{\sqrt{2}}(i|L\rangle + |R\rangle)$. These states correspond to equal superpositions with $\pm\frac{\pi}{2}$ phase shifts between the two slit components, so we expect to see a shifted pattern where the first bright band is slightly off-center. More specifically, when we detect on the left output arm of the beamsplitter, we collapse the double-slit photon state onto $|-i\rangle$, which corresponds to adding a $\frac{\pi}{2}$ phase shift to the left-slit path. This is analogous to lengthening the left-slit path, so we expect to see the interference pattern shift slightly to the left in the detection plane.

⁴We can ignore mirrors that reflect *both* beams because they would preserve the correlation or anticorrelation.

Chapter 9

Results

Because this paper is being submitted as a thesis that has a deadline, I was forced to write it while the experiment remains incomplete. A shorter, proper paper will be written upon completion of the experiment, perhaps with additional modifications to it. In this chapter, however, I discuss the amount of progress that has been made thus far, in addition to some of the challenges faced in trying to perform the experiment.

9.1 Alignment Issues

The biggest challenge faced in performing this experiment is simply getting the system properly aligned. To introduce some sarcastic terminology favored by my advisor (and his advisor before him), one might call this an “afternoon experiment”. It really shouldn’t take more than an afternoon or so to perform—that is, once its perfectly aligned.

Perhaps the trickiest alignment is placing the mirror edge such that it is correlated with the two slits of the double-slit. As visible in Figure 8.2, the mirror edge is placed on a translation stage so that its placement can be fine-tuned. Ideally, we would like the axis of translation to be parallel to the surface of the mirror, but practical constraints regarding

the mounting equipment force us to settle for a different axis. (As regards the figure, the stage has a top-to-bottom degree of freedom, transverse to the beam.) This concession makes adjustments of the mirror using the translation stage problematic, to say the least. Any adjustments necessarily alter the alignment of the reflected beam. So, our ability to optimally align the system is limited.

The physical limits of using single APD's with microscope objectives also impose alignment difficulties. One must not only ensure that the entire system is aligned up to the detection points, but that the photons are then properly coupled into the fibers that lead to the APD's. The experiment would be much easier to perform with single-photon detector arrays that can make real-time, two-dimensional detections over the entire beam profile. However, such components are beyond budget restraints.

Without these arrays, the detection takes exceedingly long amounts of time depending on what resolution we wish to obtain. This is due to the fact that we need to build up the double-slit arm pattern pixel by pixel. It is like trying to paint a pointillist piece of art, where each point takes upwards of 2 hours or more to make. Beyond simply taking time, this imposes the further restriction that the system remain stable for long periods of times—sometimes on the order of days. During this time, it is crucial that the experiment be left undisturbed. (My advisor and I have run into our fair share of issues in this regard, where despite numerous deadly warning signs, maintenance workers and others would enter the lab during an experiment. In one very recent incident, we came back after a weekend data set to find the overhead lights on in the lab. Thankfully, a dark cloth over the apparatus was enough to save our expensive APD's, which are easily saturated and destroyed by the levels of light found in a normal room.)

In order to aid in alignment, a number of apertures are used which are visible in Figure 8.2. Since the forward experiment is carried out on a single-photon level and is thus difficult to align directly, we have found it to be helpful to back-align the system with visible lasers sent

backwards through the system starting from the microscope objectives, as is typically done for single-photon experiments. For that purpose, we use a 2.5 mW, 780 nm fiber-coupled laser diode. However, even this technique can prove futile in some circumstances, as was the case with us when we were unable to align the system with any of the slits in place. Thus, it became necessary to perform a forward alignment of the system using a visible laser, which required careful introduction of the laser to the beginning of the apparatus. The difference in wavelength between our visible alignment laser and the entangled photons implies that no visible alignment will be perfect, but they are often close enough to allow easier tuning of the alignment with the single photons.

9.2 Preliminary Results

Before appealing to a forward alignment, a number of data sets had been taken, but none provided consistent significant coincidence counts with all components in. Indeed, this tragedy is what ultimately led us to resort to a visible-laser forward alignment of the system. At first the forward alignment configuration did not produce an interference pattern, which confirmed that our system had not been properly aligned. After some work, however, we were finally able to use the visible laser to form a double-slit interference pattern past the double slit. This is shown in Figure 9.1. This pattern indicates significant progress.

Although this forward alignment left us making changes that worsened our significant coincidence count rates, we took a leap of faith and ran a single photon data set without making further changes. Fortunately, we were somewhat rewarded. Figure 9.2 shows a plot of the pattern detected in coincidence as we scanned transversely behind the double-slit. In this plot, the number of random coincidences we would expect to get from the individual single-photon count rates alone is subtracted. Collecting counts for 2 hours per point, we were seeing counts of about 6 550 000 photons per time bin in the double-slit arm and about

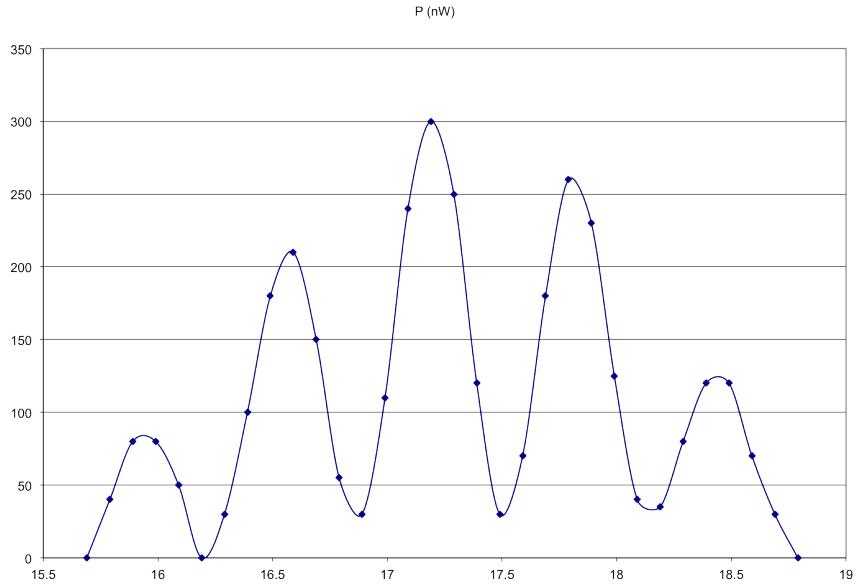


Figure 9.1: A data set demonstrating double-slit interference of a visible laser used in a forward alignment configuration. A 2.5 mW 780 nm fiber couple diode laser was passed through the apparatus depicted in Figure 8.2. The above pattern was detected past the double-slit. The vertical axis represents power in units of nW and the horizontal axis represents the transverse location of detection in units of mm.

120 000 000 in the eraser arm. This resulted in about 2700 coincidences per time bin on average. The stray negative counts in the plot indicate times when we would expect more counts from random coincidences based on our calculations. This likely is due to flukes in the actual detection time window.¹ It is clear from the graph that the pattern is far from being statistically convincing. Nevertheless, it does appear that we are seeing some interference pattern. As expected, the fringe spacing is roughly the same as the pattern we saw with the alignment beam.

Further evidence that we are seeing the eraser phenomena is shown in Figure 9.3. As

¹Postscript: Originally, we had assumed the detection time window of the APDs was 26 ns in our calculation of the expected random counts and almost half of these points were negative. Later, we used a very fast pulse generator to check this value and found we were off by 1 ns. That small change accounted for a vertical shift of over 100 counts. This discovery and subsequent correction happened after the writing of this thesis, a story for another time.

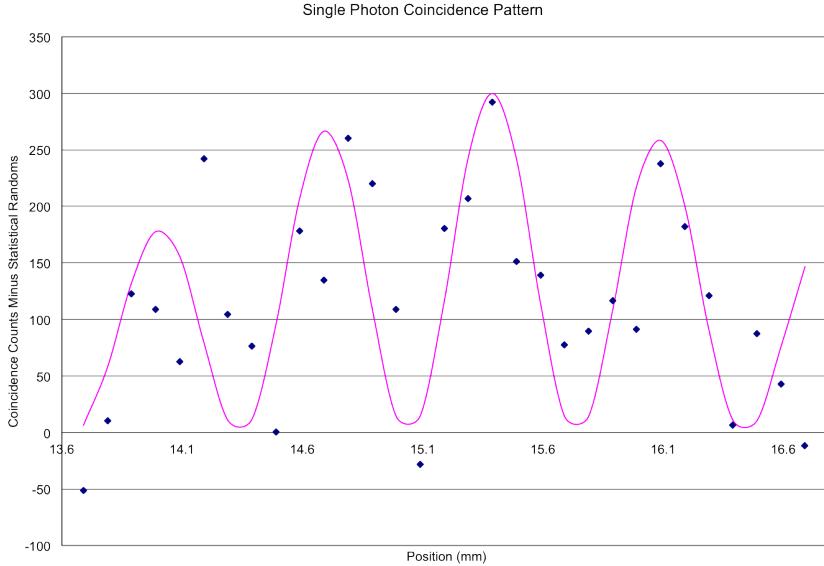


Figure 9.2: The coincidence-count pattern minus statistically expected random counts. The blue dots represent the number of coincidence counts, and the pink line is a rough fit to a product of a sinc and a cosine. The random counts are calculated statistically, based on the number of single-photon counts in each arm and the detection time window.

I discussed in Section 8.2, using a symmetric beamsplitter as an eraser would lead us to pick out a superposition state that had a $\frac{\pi}{2}$ phase shift on the left-slit component, leading to a leftwardly shifted interference pattern. Such a shift is visible in the pattern. At first it would seem that we are seeing a rightwardly shifted pattern. However, this plot is effectively taken from the perspective of someone standing at the detector and looking *back* at the double-slit. Therefore, it is proper that we see a “rightward” shift in the plot.² (I should also note that the coincidence pattern is plotted against the alignment pattern so that the same points in real space correspond to the same horizontal positions on the plot. The zero on the translation stage was shifted 2 mm between the taking of these two data sets. Thus, as visible in Figure 9.2, the coincidence pattern technically started at a point that the

²The scanning of single slit begins at the right from the perspective of someone sitting at the double slit and looking at the detector. It then scans across to the left as the data set progresses.

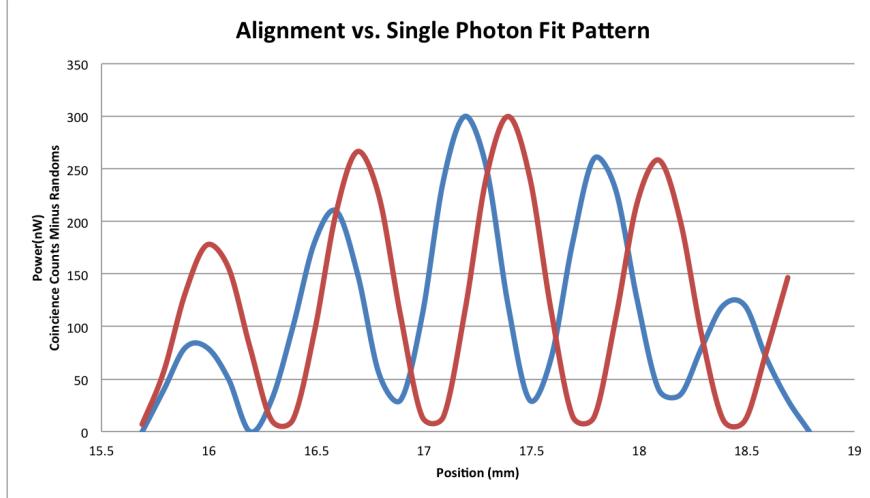


Figure 9.3: A rough fit of the coincidence pattern plotted on top of the alignment pattern. The fit of the coincidence pattern is plotted in red and the pattern from the alignment later (shown in Figure 9.1) is plotted in blue. The shift to the right of the coincidence pattern is indicative of the $\frac{\pi}{2}$ phase-shifted superposition state we would expect to generate using a symmetric beamsplitter as an eraser.

translation stage labeled 13.69 mm at the time instead of 15.69 mm.)

All of this evidence points to the fact that the eraser works. However, the data isn't entirely conclusive yet. The single-photon counts at the double slit without coincidence showed some small oscillations ranging between 6 500 000 and 6 600 000, but these seem largely insignificant. Still, our expected random background was high enough that we can't entirely rule out the possibility that the pattern in Figure 9.2 was generated by the double-slit photons alone, and that which-slit information wasn't present. As of the writing of this paper, we have begun a two-week-long data set in an effort to improve our coincidence resolution.³

³Postscript: As it turns out, this two-week-long data set did look much better and preserved all desired features. This data, or better still, will be submitted for publication as part of a more formal writeup of this work in the future.

Chapter 10

Future Work

In this chapter, I conclude this paper by briefly laying out possible avenues of future work. I begin by addressing what needs to be done with the current experiment. Then I discuss some potentially interesting modifications.

10.1 Verification of Assumptions

The most important step that must be taken in the future concerns the verification of the assumptions about the apparatus. First, we must show that our BBO is generating the momentum entanglement that we claim. Secondly, we must demonstrate that we have the proper correlation between the mirror edge and the double slit. This entails showing that eraser-arm photons can actually be used to gather good which-slit information about the double-slit photons. If a clear interference pattern is detected in our next data set, then we must also demonstrate that a π phase shift in the pattern is observed if we move the eraser-arm APD to the other output of the beamsplitter. Furthermore, the single-photon counts at the double slit should not show an interference pattern on their own. Interference should *only* be seen in the coincidence counts.

10.2 Delayed Choice

The next obvious step to take if this experiment can be successfully performed is to promote it to a delayed-choice quantum-eraser experiment. The first step would be to introduce a large delay after the detection of the double-slit photon before the eraser-arm photon reaches the eraser. This can be achieved by extending the eraser-arm photon's path by passing it through a long length of fiber optic. After this is done (and the experiment is subsequently shown to be a successful delayed-erasure quantum eraser), the next step is to introduce a method by which a choice can be made as to whether or not the erasing takes place.

One method would be to vary the length of one of the eraser-arm photon subpaths enough so that interference no longer occurs at the beamsplitter, effectively encoding which-path information in the arrival time of the eraser-arm photon (relative to the arrival time of the double-slit photon). However, it is questionable whether this could be performed quickly enough to allow for random choices to be made in the time between the detection of the double-slit photon and the arrival of the eraser-arm photon at the eraser. Another possible method would be to create a Mach–Zehnder modulator by placing a phase-modulating electro-optical modulator (EOM) at one of the outputs of the beamsplitter, then placing a second beamsplitter after this which would recombine the outputs of the first beamsplitter. The entire beamsplitter-EOM-beamsplitter sandwich could be made to act like a single beamsplitter or like nothing at all by applying different electric fields to the EOM, and this process could be switched very quickly. When the sandwich acts like a single beamsplitter it would erase which-slit information just as in the current experiment. When the sandwich acts like nothing at all we could measure which-slit information by placing APD's at the end of the two subpaths. (Which APD is triggered by the eraser-arm photon would depend on whether the eraser-arm photon was reflected by or missed the mirror edge.) The downside of this proposal is that our group does not have the extra APD it would require to com-

pletely detect this which-slit information. This, however, provides excellent motivation for our group to request the funds necessary to purchase a single-photon detector array. We could then replace the double-slit arm detector by the new array, which would not only drastically reduce the amount of time required to take data sets (*and* allow us to build up two-dimensional patterns in real time), but it would also free up an APD that could be used for measuring which-slit information.

Yet another possible method would be to use a polarization switch—a Pockels cell or polarization-modulating EOM—followed by a PBS, to quickly make decisions to redirect the eraser-arm photon into a separate sub-apparatus that would measure which-slit information. This method has the benefit that it could possibly be implemented without interfering with the alignment of the current experiment’s eraser (versus the Mach–Zehnder-modulator method which would certainly require reconfiguration and realignment of the current apparatus). However, this method would probably require another *two* APD’s. (Alternatively, we could easily measure which-slit information in the sub-apparatus using a single-photon detector array, but if we procure such an array it would be best used for detecting the double-slit photons. It is very unlikely that we would procure two arrays.) So, it is my belief that it is in our best interest to pursue the Mach–Zehnder-modulator method.

Regardless of how we decide to implement the delayed choice, we must also decide how we would ensure that the choice is made *after* the double-slit photon is detected. One option would be to constantly make choices, as fast as possible, and rely on the delay in the eraser arm in addition to some statistics to ensure that these choices usually fall in the time between the double-slit photon’s detection and the eraser-arm photon’s arrival at the eraser. Another option would be to trigger the choice on the detection of the double-slit photon. (Of course, this would also cause extra choices to be made due to false detections in the double-slit arm, but such extra choices are of no concern.)

10.3 Incoherent Pump

Another avenue of future work would be to explore what effect spatial incoherence of the pump beam has on the experiment. Most similar quantum erasers, such as the one by Kim et al. [85], use position entanglement that is merely a consequence of the spatial coherence of the pump beam across the BBO crystal. One of the novel aspects of this experiment is the use of the momentum entanglement that is a result of the intrinsic conservation of momentum during the SPDC process. Thus, it seems reasonable that we might be able to get away with a spatially incoherent source, provided that it is still collimated. However, this is far from clear, and it is very reasonable to believe such incoherence could ruin the required entangled correlations. Nevertheless, it remains an interesting consideration, both theoretically and experimentally.

Acknowledgments

This paper is the culmination of four years of study. I spent most of this time at Adelphi University pursuing my undergraduate degree in Physics and Mathematics. Additionally, I had temporary stays of research and study at Susquehanna University, the Institute for Quantum Computation (IQC) at the University of Waterloo, and the Massachusetts Institute of Technology (MIT). I would like to acknowledge all of the people—professors, students, and everyone in between—with whom I have interacted during this time. They have all shaped me in some way, and I owe very much to them. I would like to make a special acknowledgment of my thesis defense committee: Prof. Sean J. Bentley, Prof. John Dooher, and Prof. Gregory J. Mercurio. All of them have had massive impacts on my life and work. I would like to place a special emphasis on my advisor, Prof. Sean J. Bentley, without whom I would be nowhere today. I thank him not only for his support throughout this work, but for his role as my mentor throughout my entire undergraduate career. Without his influence and encouragement, I may not have chosen physics as a course of study, let alone as a career. Finally, I would like to acknowledge the family and friends who also supported me during this time, especially when I ignored them during the rushed and busy week it took me to write this paper.

Bibliography

- [1] David J. Griffiths, *Introduction to Quantum Mechanics*, 2nd ed. Pearson, 2005.
- [2] R. I. G. Hughes, *The Structure and Interpretation of Quantum Mechanics*, Fifth Printing. Harvard University Press, 1999.
- [3] Michael A. Nielsen and Isaac L. Chuang, *Quantum Computation and Quantum Information*, 10th Anniversary Edition. Cambridge University Press, 2011.
- [4] J. J. Sakurai and Jim Napolitano, *Modern Quantum Mechanics*, Second Edition. Addison-Wesley, 2011.
- [5] Linus Pauling and E. Bright Wilson, Jr., *Introduction to Quantum Mechanics with Applications to Chemistry*, Reprint. Dover, 1985.
- [6] Richard P. Feynman and Albert R. Hibbs, *Quantum Mechanics and Path Integrals*, Emended Edition. Dover, 2010.
- [7] David Griffiths, *Introduction to Elementary Particles*, Second, Revised Edition, 3rd Reprint. Wiley-VCH, 2010.
- [8] A. Zee, *Quantum Field Theory in a Nutshell*, 2nd ed. Princeton University Press, 2010.
- [9] H. J. Groenewold, Physica **12**(7), 405 (1946).

- [10] Richard P. Feynman, *The Development of the Space-Time View of Quantum Electrodynamics*. Nobel Prize Lecture, 1965.
- [11] W. Heisenberg, Zeitschrift für Physik **43**(3–4), 172 (1927).
- [12] Werner Heisenberg, *The Physical Principles of the Quantum Theory*, translated by Carl Eckart and F. C. Hoyt, reprint edition. Dover, 1949.
- [13] E. H. Kennard, Zeitschrift für Physik **44**(4–5), 362 (1927).
- [14] Hermann Weyl, *The Theory of Groups and Quantum Mechanics*, translated by H. P. Robertson, reprint edition. Dover, 1950.
- [15] H. P. Robertson, Phys. Rev. **34**(1), 163 (1929).
- [16] Y. Aharonov and D. Bohm, Phys. Rev. **122**(5), 1649 (1961).
- [17] Paul Busch, Found. Phys. **20**, 1 (1990). Also see a more recent review by Busch at arXiv:quant-ph/0105049.
- [18] Matthew Leifer, *Ten Commandments of the Church of the Smaller Hilbert Space*. <http://mattleifer.info/wordpress/wp-content/uploads/2008/11/commandments.pdf>, Accessed May 8th, 2012.
- [19] Scott Aaronson, “Better late than never”, *Shtetl-Optimized*, May 3rd, 2011. <http://www.scottaaronson.com/blog/?p=613>. Accessed May 12, 2012.
- [20] E. C. G. Stueckelberg, Helvetica Physica Acta **33**, 727 (1960).
- [21] A. Einstein, B. Podolsky, and N. Rosen, Phys. Rev. **47**, 777 (1935).
- [22] D. Bohm, *Quantum Theory*, reprint edition. Dover, 1989.
- [23] J. S. Bell, Physics **1**(3), 195 (1964).

- [24] H. P. Stapp, Phys. Rev. D **3**, 1303 (1971).
- [25] H. P. Stapp, American Journal of Physics **72**(1), 30 (2004).
- [26] P. H. Eberhard, Il Nuovo Cimento B **38**, 75 (1977).
- [27] P. H. Eberhard, Il Nuovo Cimento B **46**, 392 (1978).
- [28] P. H. Eberhard, Phys. Rev. Lett. **49**, 1474 (1982).
- [29] J. F. Clauser, M. A. Horne, A. Shimony, and R. A. Holt, Phys. Rev. Lett. **23**(15), 880 (1969).
- [30] J. F. Clauser and M. A. Horne, Phys. Rev. D **10**(2), 526 (1974).
- [31] S. J. Freedman and J. F. Clauser, Phys. Rev. Lett. **28**, 938 (1972).
- [32] G. Weihs, T. Jennewein, C. Simon, H. Weinfurter, and A. Zeilinger, Phys. Rev. Lett. **81**, 5039 (1998).
- [33] M. A. Rowe, D. Kielpinski, V. Meyer, C. A. Sackett, W. M. Itano, C. Monroe, and D. J. Wineland, Nature **409**, 791 (2001).
- [34] J. C. Howell, R. S. Bennink, S. J. Bentley, and R. W. Boyd, Phys. Rev. Lett. **92**(21), 210403 (2004).
- [35] J. F. Clauser and A. Shimony, Reports on Progress in Physics **41**, 1881 (1978).
- [36] B. S. Cirel'son, Lett. Math. Phys. **4**, 93 (1980).
- [37] A. Wilce, "Quantum Logic and Probability Theory", *The Stanford Encyclopedia of Philosophy*, Spring 2009 Edition, Edward N. Zalta (ed.). <http://plato.stanford.edu/archives/spr2009/entries/qt-quantlog/>

- [38] S. Popescu and D. Rohrlich, Found. Phys. **24**, 379 (1994); arXiv:quant-ph/9709026.
- [39] S. Kochen and E. P. Specker, Indiana Univ. Math. J. **17**(1), 59 (1968).
- [40] M. F. Pusey, J. Barrett, and T. Rudolph, Nature Physics **8**, 476 (2012). Newer version: arXiv:1111.3328 [quant-ph].
- [41] A. Peres and D. Terno, Rev. Mod. Phys. **76**, 93 (2004); arXiv:quant-ph/0212023.
- [42] J. B. Kennedy, Philosophy of Science **62**, 543 (1995).
- [43] K. A. Peacock and B. Hepburn, Proceedings of the Meeting of the Society of Exact Philosophy, (1999); arXiv:quant-ph/9906036.
- [44] P. H. Eberhard and R. R. Ross, Found. Phys. Lett. **2**(2), 127 (1989).
- [45] W. K. Wootters and W. H. Zurek, Nature **299**, 802 (1982).
- [46] D. Dieks, Phys. Lett. A **92**(6), 271 (1982).
- [47] H. Barnum, C. M. Caves, C. A. Fuchs, R. Jozsa, B. Schumacher, Phys. Rev. Lett. **76**(15), 2818 (1996); arXiv:quant-ph/9511010.
- [48] S. Lloyd, J. H. Shapiro, F. N. C. Wong, P. Kumar, S. M. Shahriar, and H. P. Yuen, Comput. Commun. Rev. **34**, 9 (2004).
- [49] Flora Lichtman, “PopSci Q&A: Seth Lloyd Talks Quantum Computing and Quoogling”, *Popular Science*, November 4th, 2011. <http://www.popsci.com/science/article/2011-10/seth-lloyd-particle-man>. Accessed May 12, 2012.
- [50] ID Quantique Press Release, *Quantum Encryption to Secure World Cup Link*, May 21st, 2010. <http://www.idquantique.com/news/press-release-worldcup.html>. Accessed May 30th, 2012.

- [51] Richard P. Feynman, Robert Leighton, and Matthew Sands, *The Feynman Lectures on Physics, Volume III*, The New Millennium Edition. Basic Books, 2011.
- [52] Richard P. Feynman, Robert Leighton, and Matthew Sands, *The Feynman Lectures on Physics, Volume I*, The New Millennium Edition. Basic Books, 2011.
- [53] A. M. March, *A Double-Slit Quantum Eraser Experiment*, <http://grad.physics.sunysb.edu/~amarch/>, Accessed May 9th, 2012.
- [54] Jordgette, *Results from the Double slit experiment: Pattern from a single slit vs. a double slit.*, http://en.wikipedia.org/wiki/File:Single_slit_and_double_slit2.jpg, Accessed May 9th, 2012.
- [55] MIT Physics Demos, *Ripple Tank: Two Point Sources*, <http://www.flickr.com/photos/physicsdemos/3745854641/in/photostream/>, <http://techtv.mit.edu/collections/physicsdemos/videos/3463-ripple-tank-interference-of-two-point-sources>, Accessed May 9th, 2012.
- [56] Max Planck, Annalen der Physik **4**(3), 553 (1901).
- [57] Albert Einstein, Annalen der Physik **17**(6), 132 (1905).
- [58] W. E. Lamb, Jr. and M. O. Scully, “The photoelectric effect without photons”, in *Polarization, Matière et Rayonnement* (Presses University de France, Paris, 1969).
- [59] G. N. Lewis, Nature **118**, 874 (1926).
- [60] T. D. Newton and E. P. Wigner, Rev. Mod. Phys. **21**, 400 (1949).
- [61] Sir Geoffrey Ingram Taylor, Proc. Camb. Phil. Soc. **15**, 114 (1909).
- [62] Claus Jönsson, Zeitschrift für Physik **161**(4), 454 (1961).

- [63] A. Tonomura, J. Endo, T. Matsuda, T. Kawasaki, and H. Ezawa, American Journal of Physics **57**(2), 117 (1989).
- [64] M. Arndt, O. Nairz, J. Vos-Andreae, C. Keller, G. van der Zouw, and A. Zeilinger, Nature **401**, 680 (1999).
- [65] T. Juffmann, A. Milic, M. Müllneritsch, P. Asenbaum, A. Tsukernik, J. Tüxen, M. Mayor, O. Cheshnovsky, and M. Arndt, Nature Nanotechnology **7**, 297 (2012).
- [66] Thomas V Marcella, arXiv:quant-ph/0703126. Earlier version: Eur. J. Phys. **23**, 615 (2002).
- [67] D. M. Greenberger and A. Yasin, Phys. Lett. A **128**(8), 391 (1988).
- [68] G. Jaeger, A. Shimony, and L. Vaidman, Phys. Rev. A **51**(1), 54 (1995).
- [69] Berthold-Georg Englert, Phys. Rev. Lett. **77**(11), 2154 (1996).
- [70] George Greenstein and Arthur G. Zajonc, *The Quantum Challenge*, Second Edition. Jones and Barlett, 2005.
- [71] R. L. Pfleegor and L. Mandel, Phys. Rev. **159**(5), 1084 (1967).
- [72] H. de Riedmatten, I. Marcikic, W. Tittel, H. Zbinden, and N. Gisin, Phys. Rev. A **67**(2), 022301 (2003).
- [73] R. Kaltenbaek, Phys. Rev. Lett. **96**(24), 240502 (2006)..
- [74] E. Schrödinger, Mathematical Proceedings of the Cambridge Philosophical Society **31**, 555 (1935).
- [75] H. M. Wiseman, S. J. Jones, and A. C. Doherty, Phys. Rev. Lett. **98**, 140402 (2007).
- [76] S. J. Jones, H. M. Wiseman, and A. C. Doherty, Phys. Rev. A **76**, 052116.

- [77] M. O. Scully, K. Drühl, Phys. Rev. **25**(4), 2208 (1982).
- [78] M. Hillery and M. O. Scully, “On State Reduction and Observation in Quantum Optics: Wigner’s Friends and Their Amnesia”, pp. 65–85 in *Quantum Optics, Experimental Gravitation, and Measurement Theory*, NATO ASI Series B, Vol. 94, P. Meystre and M. O. Scully, Eds. (Plenum, New York 1983).
- [79] M. O. Scully, B. G. Englert, and H. Walther, Nature **351**, 111 (1991).
- [80] M. O. Scully, R. Shea, and J. D. McCullen, Phys. Rep. **43**(13), 485 (1978).
- [81] P. G. Kwiat, A. M. Steinberg, and R. Y. Chiao, Phys. Rev. A **45**(11), 7729 (1992).
- [82] S. P. Walborn, M. O. Terra Cunha, S. Pádua, and C. H. Monken, Phys. Rev. A **65**, 033818 (2002).
- [83] J. A. Wheeler, p. 183 in *Mathematical Foundations of Quantum Theory*, edited by A. R. Marlow. Academic, 1978.
- [84] T. Hellmuth, H. Walther, A. Zanjonic, and W. Schleich, Phys. Rev. A. **35**(6), 2532 (1987).
- [85] Y. Kim, R. Yu, S. P. Kulik, Y. Shih, and M. O. Scully, Phys. Rev. Lett. **84**, 1 (2000).
- [86] V. Jacques, E Wu, F. Grosshans, F. Treussart, P. Grangier, A. Aspect, and J-F. Roch, Science **325**(5814), 966 (2007).