Notes on the Dual Beam Splitter Experiment

David Meyer

dmm613@gmail.com

Last Update: August 22, 2024

Initial Version: October 3, 2018

1 Introduction

This document describes an experimental setup that cannot be described in a natural way by classical physics yet still has a simple quantum explanation [1, 3]. The implication/point is this that the description of the universe given by quantum mechanics differs in fundamental ways from the classical description. In addition, this quantum description is often at odds with our intuition, which is primarily based on observations of macroscopic phenomena. Not that these observations aren't an extremely good approximation, they are, and what's more they are generally consistent with classical physics.

1.1 Preliminaries: Linear Algebra Formulation of the Circuit Model

In this section we take a brief look at the circuit model of computation. The model is described in terms of vectors and matrices. So suppose you are given a description of a circuit like the one shown in Figure 1 along with some specification of some input bit
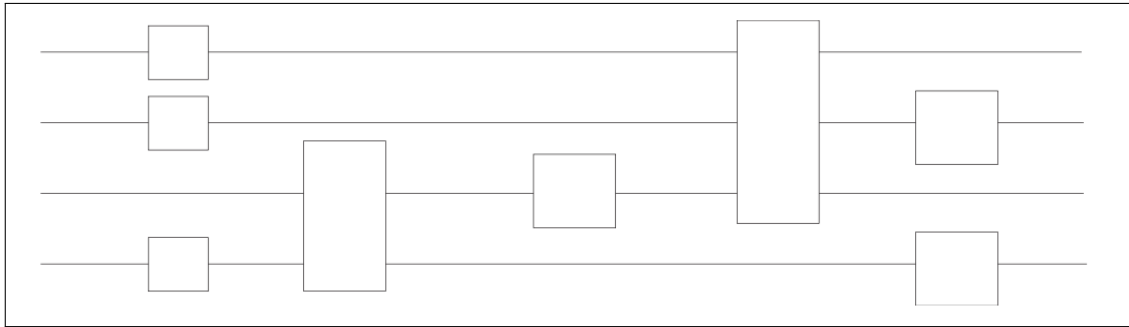

Figure 1: A curcuit of depth 5 and space 4 with 8 gates

values. If you were asked to predict the output of the circuit, you might trace through the circuit from left to right, updating the values of the bits stored on each of the wires after each gate. In other words, you would follow the *state* of the bits on the wires as they progress through the circuit. For a given point in the circuit then, we can refer to the state of the bits on the wires at that point in the circuit simply as the *state* of the computation at that point. Importantly, the *memory* of the circuit is the state of the wires, and the *gates* are transformations on that state.

Given this model, we can conceptualize the state associated with a given point in a deterministic circuit by listing the values of the bits on each of the wires in the circuit. The *state* of any particular wire at a given point in a circuit, of course, is just the value of the bit on that wire (0 or 1). For a probabilistic circuit, however, this simple description is not enough. In particular, consider a single bit that is in state 0 with probability p_0 and in state 1 with probability p_1 (where $p_0 + p_1 = 1$). We can represent this state by a 2-dimensional vector of probabilities:

$$\begin{pmatrix} p_0 \\ p_1 \end{pmatrix}$$

Note that this description can also be used for deterministic circuits. A wire in a deterministic circuit whose state is 0 could be specified by the probabilities $p_0 = 1$ and $p_1 = 0$, and the corresponding vector

$$\begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

Similarly, a wire in state 1 could be represented by the probabilities $p_0 = 0$, $p_1 = 1$, and the vector

$$\begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

Since the standard representation the states of wires (or collections of wires) in a circuit is by vectors, we will also want to represent gates in the circuit by operators that act on these state vectors. Hence these operators are most conveniently described by matrices. Consider, for example, the logical NOT gate. We would like to define an operator (matrix) that behaves on state vectors in a manner consistent with the behavior of the NOT gate. If we know a wire is in state 0 (so $p_0 = 1$), the not gate maps it to state 1 (so $p_1 = 1$), and vice versa. In terms of the vector representations of these states, we have

$$\text{NOT} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

Similarly,

$$\text{NOT} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

The next step is to characterize the *transformation* (gate) as a matrix. In the case of the NOT gate, we have

$$\text{NOT} \equiv \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

So now if we want to apply the NOT gate to a wire in a given state, we multiply the state vector on the left by the matrix representation of the NOT gate:

$$\text{NOT} \begin{pmatrix} p_0 \\ p_1 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} p_0 \\ p_1 \end{pmatrix}$$

Now we want to describe the state associated with a given point in a probabilistic circuit having two wires. Let the state of the first wire at the given point be 0 with probability p_0 and 1 with probability p_1 and the state of the second wire at this point be 0 with probability q_0 and 1 with probability q_1. The four possibilities for the combined state of both wires at the given point are $\{00,01,10,11\}$, where the binary string ij indicates that the first wire is in state i and the second wire in state j. Then the combined state of both wires can be described by the column vector of probabilities:

$$\begin{pmatrix} p_0 q_0 \\ p_0 q_1 \\ p_1 q_0 \\ p_1 q_1 \end{pmatrix}$$

What this is setting up is that the state is really the *tensor product* of the 2-dimensional vectors for the states of the first and second wires:

$$\begin{pmatrix} p_0 q_0 \\ p_0 q_1 \\ p_1 q_0 \\ p_1 q_1 \end{pmatrix} = \begin{pmatrix} p_0 \\ p_1 \end{pmatrix} \otimes \begin{pmatrix} q_0 \\ q_1 \end{pmatrix}$$

1.2 Aside: The Dirac Notation

The Dirac Notation, also known as bra-ket notation, was invented by Paul Dirac for use in quantum mechanics [2]. The problem Dirac faced was that, while column vector notation is ubiquitous in linear algebra, the standard notation was often cumbersome in quantum mechanical derivations. This is especially true when dealing with both inner and outer products (e.g., with multiple qubits). For example, when we define ψ to be a vector it is not clear whether ψ is a row or a column vector. Similarly, if ϕ and ψ are vectors then it is equally unclear if $\phi\psi$ is even defined because the shapes of ϕ and ψ may be unclear in the context[1]. Beyond the ambiguity about the shapes of vectors, expressing even simple vectors using the linear algebraic notation introduced earlier can be very less elegant. For example, if we wish to describe an n-qubit state where each qubit takes the value 0 then we would express the state as

$$\begin{pmatrix} 1 \\ 0 \end{pmatrix} \otimes \ldots \otimes \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

So one of the problems here is that evaluating this tensor product is impractical because the vector lies in an exponentially large space. So maybe this notation is the best description of the state that can be given using the previous notation (it is really speculation; I have no idea how one would prove such an assertion). In any event, Dirac's approach was to define a new notation that is designed to fit the needs of quantum mechanics.

1.2.1 Preliminaries

In order to confront both the bulky notation of linear algebra (at least when applied to quantum mechanics use cases) and the ambiguity posed by standard vector notation, Dirac proposed a new notation for two kinds of vectors frequently used in quantum mechanics: a "ket" vector, denoted $|a\rangle$, and a "bra" vector, which is denoted $\langle a|$. The naming comes from the fact that when put together they form a "bra[c]ket" or inner product. So if If ϕ is a column vector then we can write it in Dirac notation as $|\phi\rangle$, where the $|\cdot\rangle$ denotes that it is a unit column vector, i.e., a ket vector. Similarly, the row vector $\psi^\dagger$ is expressed as $\langle\psi|$, where $\psi^\dagger$ is obtained by applying entry-wise complex conjugation to the elements of the transpose of ψ (more on this below). The bra-ket notation also implies that $\langle\psi|\psi\rangle$ is the inner product of vector ψ with itself, which is by definition 1.

The ket vector can be represented as a column vector:

---

[1]Recall that matrix multiplication is only defined when the number of columns of the first matrix (vector) is equal to the number of rows in the second matrix.
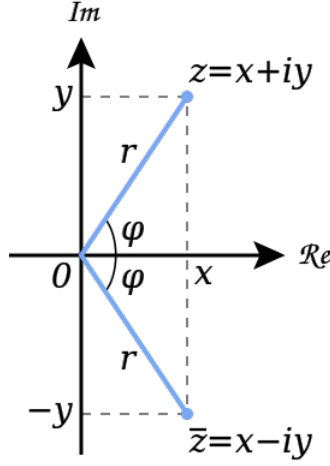
Figure 2: Complex conjugates. Image courtesy Wikipedia.

$$|a\rangle = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}$$

where n is typically three or four (representing the dimensionality of space or space-time) and a_i is a complex number, namely $a_i \in \mathbb{C}$.

Similarly, a bra vector can be represented as

$$\langle a| = \begin{pmatrix} a_1^* & a_2^* & \ldots & a_n^* \end{pmatrix}$$

where a_i^* is the *complex conjugate* of a_i. That is, if z is a complex number, it can be written as $z = x + iy$, where $x, y \in \mathbb{R}$ and $i = \sqrt{-1}$. Then the complex conjugate of z, denoted z^* where $z^* = x - iy$, is the reflection of z around the real axis. This is depicted in Figure 2 (note that $\bar{z} = z^*$ in Figure 2).

The *modulus* of z, denoted as $|z|^2$, is the length of the directed line segment drawn from the origin of the complex plane to the point (x, y). This directed line segment is also the vector that represented by $x + iy$, so the modulus of a complex number is the same thing as the magnitude (or length) of the vector representing z. As a result $|z|^2 = x^2 + y^2$ by the Pythagorean Theorem. This also means that $|z|^2 = |z^*|^2$.

Another interesting property of z is that if you multiply z by its complex conjugate z^* you get

$$zz^* = (x + iy)(x - iy)$$
$$= x^2 + y^2$$

Note also that the complex conjugate of some $x \in \mathbb{R}$ is simply x. That is, for $x \in \mathbb{R}$, $x = x^*$.

Now we're ready to do some arithmetic. Consider first that $|a\rangle + |b\rangle = |c\rangle$. Why? We can see this (without loss of generality) in three dimensions:

$$\begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} + \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} = \begin{pmatrix} a_1 + b_1 \\ a_2 + b_2 \\ a_3 + b_3 \end{pmatrix}$$

ket vectors, being column vectors, follow basic linear algebra properties, such as

$$\alpha \, |a\rangle = \alpha \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} = \begin{pmatrix} \alpha a_1 \\ \alpha a_2 \\ \alpha a_3 \end{pmatrix}$$

Now, what happens if you multiply a bra vector by a ket vector? That is, $\langle a|b \rangle$? Here the brackets indicate the inner product. In our three dimensional example, we get

$$\langle a|b \rangle = \begin{pmatrix} a_1^* & a_2^* & a_3^* \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} = a_1^* b_1 + a_2^* b_2 + a_3^* b_3$$

Similarly, $\langle b|a \rangle = b_1^* a_1 + b_2^* a_2 + b_3^* a_3$. What about $\langle b|a \rangle^*$? Well, since $\langle b|a \rangle^* = b_1 a_1^* + b_2 a_2^* + b_3 a_3^*$ and multiplication is commutative in $\mathbb{C}$ we see that $\langle b|a \rangle^* = \langle a|b \rangle$.

Now lets look at operations on a square matrix M. Suppose that M is a 3×3 matrix. Then we can write M as the matrix of elements $m_{i,j}$, where $m_{i,j}$ is the element in the ith row and j^{th} column, as follows

$$M = \begin{pmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{31} & m_{32} & m_{33} \end{pmatrix}$$

What about M^*? Well this is just

$$M^* = \begin{pmatrix} m_{11}^* & m_{12}^* & m_{13}^* \\ m_{21}^* & m_{22}^* & m_{23}^* \\ m_{31}^* & m_{32}^* & m_{33}^* \end{pmatrix}$$

The *transpose* of a matrix M, M^{T} is defined as

$$M^{\mathrm{T}} = \begin{pmatrix} m_{11} & m_{21} & m_{31} \\ m_{12} & m_{22} & m_{32} \\ m_{13} & m_{23} & m_{33} \end{pmatrix}$$

Essentially the transpose of a matrix is the reflection around the diagonal. Another important matrix, $M^\dagger$, is the combination of the transposition and the complex conjugate. That is

$$M^\dagger = \begin{pmatrix} m_{11}^* & m_{21}^* & m_{31}^* \\ m_{12}^* & m_{22}^* & m_{32}^* \\ m_{13}^* & m_{23}^* & m_{33}^* \end{pmatrix}$$

In the case where $M = M^\dagger$, M is called Hermitian. For a (somewhat contrived) example suppose

$$M = \begin{pmatrix} 1 & i+1 \\ i-1 & 3 \end{pmatrix}$$

To find $M^\dagger$, we first take the transpose of M

$$M^{\mathrm{T}} = \begin{pmatrix} 1 & i-1 \\ i+1 & 3 \end{pmatrix}$$

and then take the complex conjugate, resulting in

$$M^\dagger = \begin{pmatrix} 1 & i+1 \\ i-1 & 3 \end{pmatrix}$$

or just M. So we can see that in this case, $M = M^\dagger$, that is, M is Hermitian.

Next we want to ask what happens if we have a matrix times a ket vector, that is, $M\,|a\rangle$. An example in two dimensions would look like

$$M\,|a\rangle = \begin{pmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} m_{11}a_1 + m_{12}a_2 \\ m_{21}a_1 + m_{22}a_2 \end{pmatrix} = |c\rangle$$

But what is $|c\rangle$? We need to define a few other matrices in order to answer this question.

First, the Identity matrix I (three dimensional in this case) is

$$I = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Another important matrix, *Unitary* matrix U is a matrix that has the interesting property that $UU^\dagger = I$. U is a matrix such that $U^\dagger$ is the inverse of U. A few other operations will become important later. These include the *determinant* of M in the 2×2 matrix case is

$$\text{Det } M = \begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc$$

$$\text{Det } I = \begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix} = 1 \cdot 1 - 0 \cdot 0 = 1$$

The *Trace* of a matrix Tr M is defined as the sum of the elements on the main diagonal. So if for example M is two dimensional

$$\text{Tr } M = \text{Tr} \begin{pmatrix} a & b \\ c & d \end{pmatrix} = a + d$$

Back to the case where M is a Hermitian matrix[2]. In this case we call the matrix H (instead of M). So what is $H\,|a\rangle$? Well, kind of amazingly

$$H\,|a\rangle = \lambda\,|a\rangle$$

---

[2]If $M = M^\dagger$, M is Hermitian.

where $|a\rangle$ is an *eigenvector* and λ is an *eigenvalue*. In words: A Hermitian matrix acting on a ket vector produces the same ket vector scaled by a real number $\lambda \in \mathbb{R}$.
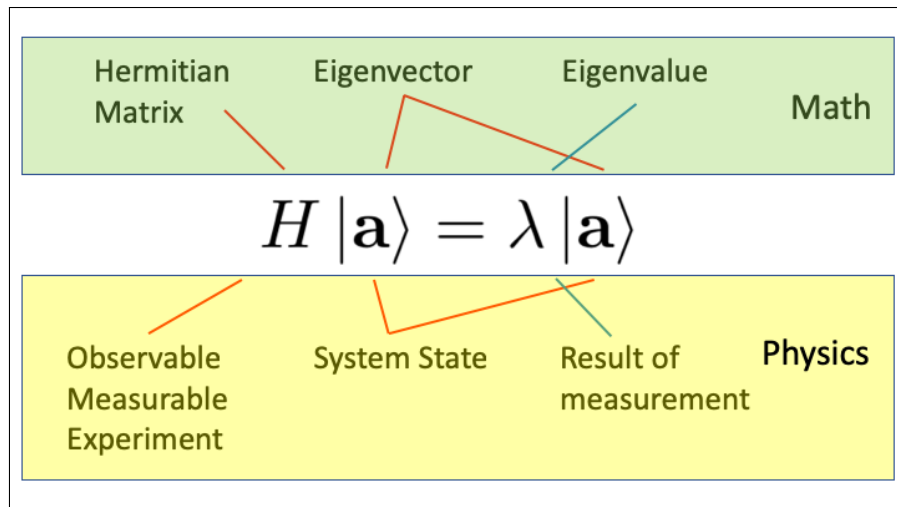

Figure 3: Mathematical and physical interpretations of $H|a\rangle = \lambda|a\rangle$

Looking ahead, the Hermitian matrix (operator) H will represent something that is *observable* or *measurable*, i.e., an *experiment*, the eigenvector $|a\rangle$ represents the *state* of the system and the eigenvalue $\lambda \in \mathbb{R}$ represents the actual outcome of the measurements. This is illustrated (I hope) in Figure 3.

1.3   Review: Orthonormality, Completeness, and Projection

As we saw above, unitary matrices are matrices which satisfy

$$\mathbf{U}^{-1} = \mathbf{U}^{\dagger} \tag{1}$$

Unitary matricies are ubiquitous and important in quantum mechanics, in particular because they have the following unique and useful properties: Orthonormality, Completeness, and Projection. We'll briefly look at each of these below[3].

1.3.1   Orthornomality

We can rewrite Equation 1 as

---

[3]I will use the notation $\begin{pmatrix} x_1 & \dots & x_n \end{pmatrix}^{\mathrm{T}}$ and $\begin{bmatrix} x_1 & \dots & x_n \end{bmatrix}^{\mathrm{T}}$ interchangably in the following discussion.

$$\mathbf{U}^\dagger \mathbf{U} = \mathbf{I} \tag{2}$$

where $\mathbf{I}$ is the identity matrix. What Equation 2 is really telling us is that the columns of the matrix $\mathbf{U}$ form a set of orthnormal vectors.

Note that we can interpret a matrix as a row vector where the entries are the columns $\mathbf{v}_i$ of $\mathbf{U}$. That is

$$\mathbf{U} = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \ldots & \mathbf{v}_N \end{bmatrix}$$

Similarly, $\mathbf{U}^{-1}$ can be written as a column vector where the entries are the row vectors $\mathbf{v}_i^\dagger$:

$$\mathbf{U}^{-1} = \mathbf{U}^\dagger = \begin{bmatrix} \mathbf{v}_1^\dagger \\ \mathbf{v}_2^\dagger \\ \vdots \\ \mathbf{v}_N^\dagger \end{bmatrix}$$

Now we can see that

$$
\begin{aligned}
\mathbf{U}^\dagger \mathbf{U} &= \begin{bmatrix} \mathbf{v}_1^\dagger \\ \mathbf{v}_2^\dagger \\ \vdots \\ \mathbf{v}_N^\dagger \end{bmatrix} \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \ldots & \mathbf{v}_N \end{bmatrix} \\
&= \begin{bmatrix}
\mathbf{v}_1^\dagger \cdot \mathbf{v}_1 & \mathbf{v}_1^\dagger \cdot \mathbf{v}_2 & \mathbf{v}_1^\dagger \cdot \mathbf{v}_3 & \ldots & \mathbf{v}_1^\dagger \cdot \mathbf{v}_N \\
\mathbf{v}_2^\dagger \cdot \mathbf{v}_1 & \mathbf{v}_2^\dagger \cdot \mathbf{v}_2 & \mathbf{v}_2^\dagger \cdot \mathbf{v}_3 & \ldots & \mathbf{v}_2^\dagger \cdot \mathbf{v}_N \\
\mathbf{v}_3^\dagger \cdot \mathbf{v}_1 & \mathbf{v}_3^\dagger \cdot \mathbf{v}_2 & \mathbf{v}_3^\dagger \cdot \mathbf{v}_3 & \ldots & \mathbf{v}_3^\dagger \cdot \mathbf{v}_N \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
\mathbf{v}_N^\dagger \cdot \mathbf{v}_1 & \mathbf{v}_N^\dagger \cdot \mathbf{v}_2 & \mathbf{v}_N^\dagger \cdot \mathbf{v}_3 & \ldots & \mathbf{v}_N^\dagger \cdot \mathbf{v}_N
\end{bmatrix} \\
&= \mathbf{I}
\end{aligned}
$$

or in Dirac notation

$$\mathbf{U}^\dagger\mathbf{U} = \begin{bmatrix} \mathbf{v}_1^\dagger \\ \mathbf{v}_2^\dagger \\ \vdots \\ \mathbf{v}_N^\dagger \end{bmatrix} \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \dots & \mathbf{v}_N \end{bmatrix}$$

$$= \begin{bmatrix} \langle v_1| \\ \langle v_2| \\ \vdots \\ \langle v_N| \end{bmatrix} \begin{bmatrix} |v_1\rangle & |v_2\rangle & \dots & |v_N\rangle \end{bmatrix}$$

$$= \begin{bmatrix} \langle v_1|v_1\rangle & \langle v_1|v_2\rangle & \langle v_1|v_3\rangle & \dots & \langle v_1|v_N\rangle \\ \langle v_2|v_1\rangle & \langle v_2|v_2\rangle & \langle v_2|v_3\rangle & \dots & \langle v_2|v_N\rangle \\ \langle v_3|v_1\rangle & \langle v_3|v_2\rangle & \langle v_3|v_3\rangle & \dots & \langle v_3|v_N\rangle \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \langle v_N|v_1\rangle & \langle v_N|v_2\rangle & \langle v_N|v_3\rangle & \dots & \langle v_N|v_N\rangle \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{bmatrix}$$

$$= \mathbf{I}$$

Note here that $\langle v_i|v_i\rangle = 1$ (the v_i are unit vectors) and $\langle v_i|v_j\rangle = 0$ for $i \neq j$ (v_i and v_j are orthogonal). In quantum mechanics two states v_i and v_j are said to be distinguishable or measurable if they are orthogonal, that is, if $\langle v_i|v_j\rangle = 0$.

Another way to say this to notice[4] that since $(\mathbf{U}^\dagger\mathbf{U})_{ij} = (\mathbf{U}^{-1}\mathbf{U})_{ij} = \delta_{ij}$, the columns of $\mathbf{U}$ can be written as the inner product $\langle v_i|v_j\rangle = \delta_{ij}$. Said another way, the vectors v_i form an orthonormal set. In particular, if $\mathbf{V} = \{v_j\}$ is an orthonormal set, then for $v_i, v_j \in \mathbf{V}$, the inner product $\langle v_i|v_j\rangle = \delta_{ij}$.

1.3.2 Completeness

From $\mathbf{U}^\dagger\mathbf{U} = \mathbf{I}$ we saw that we could derive orthonormality. But we also expect that $\mathbf{U}\mathbf{U}^\dagger = \mathbf{I}$. It turns out that we can get something interesting by observing this. In particular

---

[4] δ_{ij} is the Kronecker Delta function [4], $\delta_{ij} = \begin{cases} 1 & \text{when } i = j \\ 0 & \text{when } i \neq j \end{cases}$

$$\mathbf{U}\mathbf{U}^\dagger = \begin{bmatrix} |v_1\rangle & |v_2\rangle & |v_3\rangle & \dots & |v_N\rangle \end{bmatrix} \begin{bmatrix} \langle v_1| \\ \langle v_2| \\ \langle v_3| \\ \vdots \\ \langle v_N| \end{bmatrix} = \mathbf{I}$$

If we multiply this out, we find that

$$|v_1\rangle \langle v_1| + |v_2\rangle \langle v_2| + \dots + |v_N\rangle \langle v_N| = \sum_{i=1}^{N} |v_i\rangle \langle v_i| = \mathbf{I} \tag{3}$$

Equation 3 is known as the *completeness* relation. Completeness turns out to be useful and is a sort of flip-side of orthonormality. While orthonormality is kind of an "inner product" ($\mathbf{U}^\dagger\mathbf{U}$), completeness is like an outer product in that $\mathbf{U}\mathbf{U}^\dagger$ is a sum over i of $|v_i\rangle \langle v_i|$. Interestingly, the trace of the outer product of two $n \times 1$ column vectors $\mathbf{a}$ and $\mathbf{b}$ is $\text{Tr}(|a\rangle \langle b|) = \langle a|b\rangle$.

1.3.3 Projection

To get an idea of what projection is all about, consider the expansion of a vector into components in a basis:

$$|w\rangle = \sum_{i=1}^{N} w_i |v_i\rangle \tag{4}$$

Now, if the set of vectors basis vectors $\{v_i\}$ are orthonormal, then we know that

$$w_i = \langle v_i|w\rangle$$

and substituting back into Equation 4 we get

$$|w\rangle = \sum_{i=1}^{N} \langle v_i|w\rangle |v_i\rangle$$

Interestingly, there is another way to derive this result: use the completeness relation, which is simply a fancy but useful way to write $\mathbf{I}$:

$$|w\rangle = \mathbf{I} \cdot |w\rangle = \left(\sum_{i=1}^{N} |v_i\rangle \langle v_i| \right) |w\rangle = \sum_{i=1}^{N} |v_i\rangle \langle v_i|w\rangle$$

In words, we were able to use the completeness relation to project a vector onto its components in a particular basis.

For example, we know that for vectors $|\alpha\rangle$ and $|\beta\rangle$, we can take the inner product between them by using their components in a basis $\{v_i\}$:

$$\langle \alpha|\beta \rangle = \sum_{i=1}^{N} a_i^* b_i$$

where $a_i = \langle v_i|\alpha \rangle$ and $b_i = \langle v_i|\beta \rangle$. Interestingly, we can again derive this using the completeness relation:

$$
\begin{aligned}
\langle \alpha|\beta \rangle &= \langle \alpha| \mathbf{I} |\beta \rangle & &\# \langle \alpha|\beta \rangle = \langle \alpha| \, |\beta \rangle = \langle \alpha| \mathbf{I} |\beta \rangle \\
&= \langle \alpha| \left(\sum_{i=1}^{N} |v_i\rangle \langle v_i| \right) |\beta \rangle & &\# \sum_{i=1}^{N} |v_i\rangle \langle v_i| = \mathbf{I} \ \ \text{(Equation 3)} \\
&= \sum_{i=1}^{N} \langle \alpha|v_i \rangle \langle v_i|\beta \rangle & &\# \ \text{rearrange} \\
&= \sum_{i=1}^{N} \langle v_i|\alpha \rangle^* \langle v_i|\beta \rangle & &\# \langle a|b \rangle = \langle b|a \rangle^* \ \text{so} \ \langle \alpha|v_i \rangle = \langle v_i|\alpha \rangle^* \\
&= \sum_{i=1}^{N} a_i^* b_i & &\# \ a_i^* = \langle v_i|\alpha \rangle^* \ \text{and} \ b_i = \langle v_i|\beta \rangle
\end{aligned}
$$

13

1.4 Photon Polarization Examples
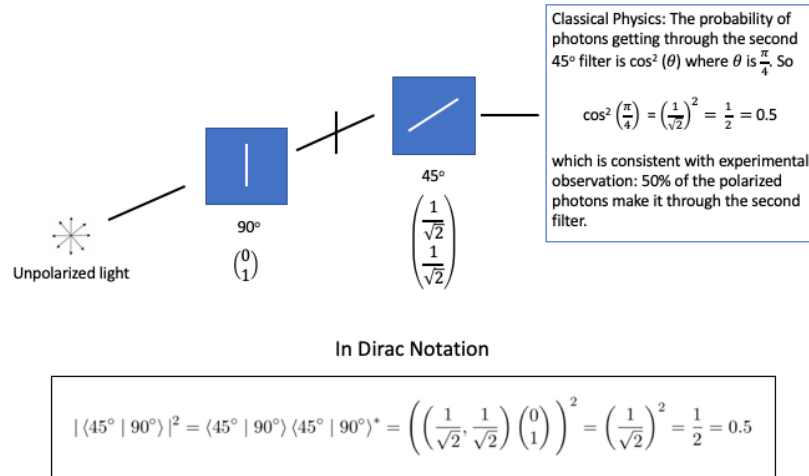
We know from experimental evidence that ... TBD

Classical Physics: The probability of photons getting through the second 45° filter is $\cos^2(\theta)$ where θ is $\frac{\pi}{4}$. So

$$\cos^2\left(\frac{\pi}{4}\right) = \left(\frac{1}{\sqrt{2}}\right)^2 = \frac{1}{2} = 0.5$$

which is consistent with experimental observation: 50% of the polarized photons make it through the second filter.

In Dirac Notation

$$| \langle 45° \mid 90° \rangle |^2 = \langle 45° \mid 90° \rangle \langle 45° \mid 90° \rangle^* = \left(\left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}} \right) \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right)^2 = \left(\frac{1}{\sqrt{2}} \right)^2 = \frac{1}{2} = 0.5$$

Figure 4: What is the probability of a vertically polarized photon getting through a 45° polarizer?

$$| \langle 45° \mid 90° \rangle |^2 = \langle 45° \mid 90° \rangle \langle 45° \mid 90° \rangle^* = \left(\left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}} \right) \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right)^2 = \left(\frac{1}{\sqrt{2}} \right)^2 = \frac{1}{2} = 0.5$$

Here

$$| \theta \rangle = \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \alpha \, |x\rangle + \beta \, |y\rangle$$

where

$$\alpha = \cos(\theta)$$
$$\beta = \sin(\theta)$$

and the *basis vectors* $|x\rangle$ and $|y\rangle$ are

$$|x\rangle = \begin{pmatrix} \cos(0) \\ \sin(0) \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

$$|y\rangle = \begin{pmatrix} \cos(\frac{\pi}{2}) \\ \sin(\frac{\pi}{2}) \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

These basis vectors can be thought of as a vector $\begin{pmatrix} x \\ y \end{pmatrix}$ where x is the projection of a vector r on to the x axis and y is the projection of r on the y axis. This is depicted in Figure 5.
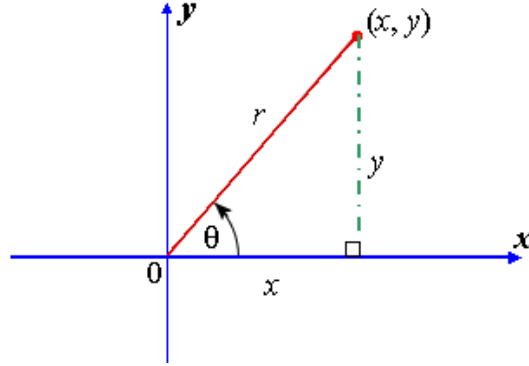

Figure 5: Two Dimensional Space

Back to our example: here the matrix representation of $(\alpha_0 |0\rangle + \alpha_1 |1\rangle) + (\beta_0 |0\rangle + \beta_1 |1\rangle)$ is the tensor product

$$\begin{pmatrix} \alpha_0 \\ \alpha_1 \end{pmatrix} \otimes \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} = \begin{pmatrix} \alpha_0\beta_0 \\ \alpha_0\beta_1 \\ \alpha_1\beta_0 \\ \alpha_1\beta_1 \end{pmatrix}$$

1.5 The *Hadamard* Transform

As mentioned above, a unitary matrix that acts on a small number of qubits is often called a gate, in analogy to classical logic gates like AND, OR, and NOT. Two simple but important 1-qubit gates are the bit-flip gate X (which negates the bit, i.e., swaps $|0\rangle$ and $|1\rangle$) and the phase-flip gate Z (which puts a $-$ in front of $|1\rangle$). The 2×2 unitary matrix representations of X and Z are

$$X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

15

Another important 1-qubit gate is the *phase gate* R_ϕ, which rotates the phase of the $|1\rangle$-state by an angle ϕ such that

$$R_\phi |0\rangle = |0\rangle$$
$$R_\phi |1\rangle = e^{i\phi} |1\rangle$$

This corresponds to the unitary matrix

$$R_\phi = \begin{pmatrix} 1 & 0 \\ 0 & e^{i\phi} \end{pmatrix}$$

Interestingly Z is a special case of R_ϕ: $Z = R_\pi$ since by Euler's identity $e^{i\pi} = -1$.

Possibly the most important 1-qubit gate is the *Hadamard transform*, specified by:

$$H |0\rangle = \frac{1}{\sqrt{2}} |0\rangle + \frac{1}{\sqrt{2}} |1\rangle$$
$$H |1\rangle = \frac{1}{\sqrt{2}} |0\rangle - \frac{1}{\sqrt{2}} |1\rangle$$

The (unitary) matrix representation of *Hadamard transform* is then

$$H = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 0 & -1 \end{pmatrix}$$

If we apply H to initial state $|0\rangle$ and then measure, we have equal probability of observing $|0\rangle$ or $|1\rangle$. Similarly, applying H to $|1\rangle$ and observing gives equal probability of $|0\rangle$ or $|1\rangle$.

However, if we apply H to the superposition $\frac{1}{\sqrt{2}} |0\rangle + \frac{1}{\sqrt{2}} |1\rangle$ we get $|0\rangle$, since

$$H |0\rangle = \frac{1}{\sqrt{2}} |0\rangle + \frac{1}{\sqrt{2}} |1\rangle$$
$$H |1\rangle = \frac{1}{\sqrt{2}} |0\rangle - \frac{1}{\sqrt{2}} |1\rangle$$
$$H(\frac{1}{\sqrt{2}} |0\rangle + \frac{1}{\sqrt{2}} |1\rangle) = \frac{1}{2}(|0\rangle) + |1\rangle) + \frac{1}{2}(|0\rangle) - |1\rangle) = |0\rangle$$
$$H(\frac{1}{\sqrt{2}} |0\rangle - \frac{1}{\sqrt{2}} |1\rangle) = \frac{1}{2}(|0\rangle) + |1\rangle) - \frac{1}{2}(|0\rangle) - |1\rangle) = |1\rangle$$

16

What has happened here is that the positive and negative amplitudes for $|1\rangle$ have canceled out[5]. This effect is called *interference*, and is analogous to interference patterns between light or sound waves. We will see this again in Section 3.

Finally, notice that one application of the Hadamard gate to either a 0 or 1 qubit will produce a quantum state that, if observed, will be a 0 or 1 with equal probability. This is exactly like flipping a fair coin in the standard probabilistic model of computation. However, if the Hadamard gate is applied twice in succession, then the final state is always the same as the initial state. This would be like taking a fair coin that is showing heads, flipping it twice, and it always landing on heads after the second flip.

1.6   One More Example: CNOT

We can also represent gates acting on more than one wire. For example, consider the controlled-not gate, denoted CNOT. This is a gate that acts on two bits, labelled the control bit (c) and the target bit (t). The action of the gate is to apply the not operation to the target if the control bit is 0, and do nothing otherwise (the control bit is always unaffected by the CNOT gate). Equivalently, if the state of the control bit is c, and the target bit is in state t the CNOT gate maps the target bit to $t \oplus c$ (where $\oplus$ represents the logical exclusive-or operation, or addition modulo 2). The CNOT gate is illustrated in Figure 6.

$$\text{CNOT} \equiv \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}.$$
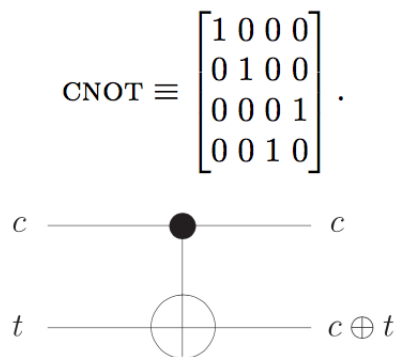

Figure 6: The *reversible* CNOT gate flips the value of the t iff the control bit $c = 1$

Consider a pair of wires such that the first wire is in state 1 and the second in state 0. This means that the 4-dimensional vector describing the combined state of the pair of wires is

---

[5]This also means that H is its own inverse.

$$\begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} \tag{5}$$

We now want to apply the CNOT gate to a pair of wires shown in Equation 5 where we interpret the first wire as the control bit and the second as the target bit. From the description of the CNOT gate, we expect the result should be that the control bit (first wire) remains in state 1, and the target bit (second wire) flips to state 1. That is, we expect the resulting state vector to be

$$\begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}$$

That is

$$\text{CNOT} \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} \equiv \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}$$

2 Beam Splitters

Suppose we have an experimental setup consisting of a photon source, a beam splitter (which was once implemented using a half-silvered mirror), and a pair of photon detectors. This is the classic beam splitter setup, and is illustrated in Figure 7.

Suppose we send a series of individual photons along a path from the photon source towards the beam splitter. We observe the photon arriving at the detector on the right on the beam splitter half of the time, and arriving at the detector above the beam splitter half of the time, as illustrated in Figure 8. Perhaps the simplest way to explain this behavior in a theory of physics is to model the beam splitter as effectively flipping a fair coin, and choosing whether to transmit or reflect the photon based on the result of the coin-flip, whose outcome determines whether the photon is transmitted or reflected. This is very similar to the behavior of a probabilistic Turing machine.

Now consider a modification of the setup shown in Figure 7 with a pair of beam splitters and fully reflecting mirrors to direct the photons along either of two paths as shown in
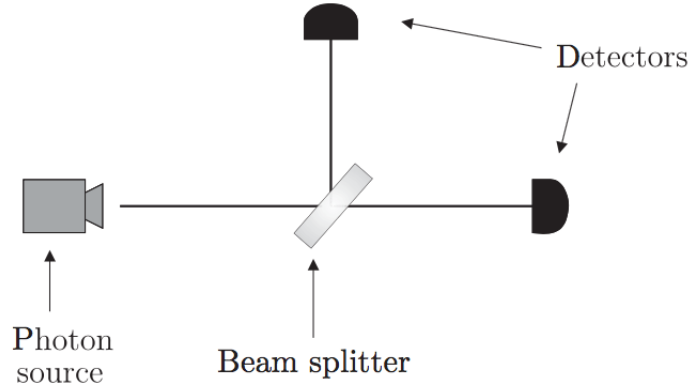
Figure 7: Experimental setup with one beam splitter

Figure 9, where the paths are labelled 0 and 1. This setup is the famous the Mach-Zehnder interferometer [5] It is important to note that the length of paths 0 and 1 are equal, so the photons arrive at the same time, regardless of which path is taken.

By treating the beam splitters as independently deciding at random whether to transmit or reflect incident photons, classical physics predicts that each of the detectors will register photons arriving 50 per cent of the time, on average. Here, however, the results of experiments reveal an entirely different behavior. However, what we find is that the photons arrive at only one of the detectors, 100 per cent of the time. This is shown in Figure 10. But why?

The result of the experiment shown in Figure 10 is at the very least counter-intuitive, largely because it does not agree with our classical intuition. However, quantum physics models the experiment in a way that correctly predicts the observed outcomes. The non-intuitive behavior results from features of quantum mechanics called *superposition* and *interference*. Let's take a look at a framework that can explain the results of this experiment.

First, let's suppose that the second beam splitter were not present in the apparatus. Then the photon follows one of two paths (according to classical physics), depending on whether it is reflected or transmitted by the first beam splitter. If it is transmitted through the first beam splitter, the photon arrives at the top detector, and if it is reflected, the photon arrives at the detector on the right. We can consider a photon in the apparatus as a 2-state system, letting the presence of the photon in one path represent a 0 and letting the presence of the photon in the other path represent a 1.

Following our approach with classical circuits, we denote the state of a photon in path 0
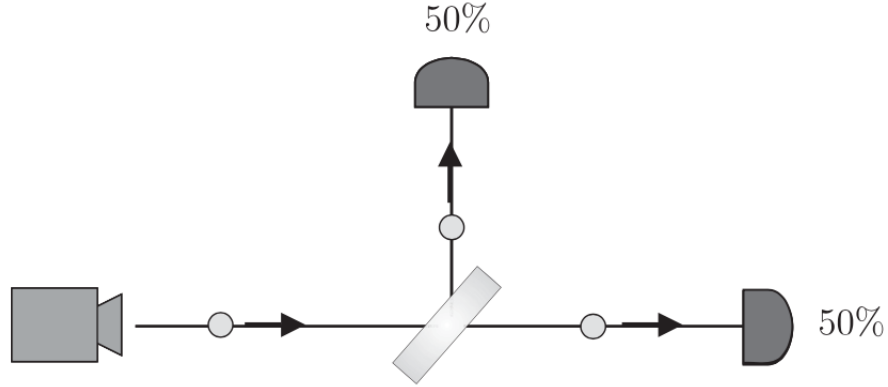
Figure 8: Experimental setup with one beam splitter

by the vector

$$\begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

and a photon in path 1 by the vector

$$\begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

By convention, the photon leaving the source starts out in the 0 path. In a classical description, we model the beam splitter as randomly selecting whether the photon will continue along the 0 path, or be reflected into the 1 path. According to the quantum mechanical description, the beam splitter causes the photon to go into a superposition of taking both the 0 and 1 paths. Mathematically, we describe such a superposition by taking a linear combination of the state vectors for the 0 and 1 paths, so the general path state will be described by the vector shown in Equation 2.

If we were to physically measure the photon to see which path it is in, we will find it in path 0 with probability $|\alpha_0|^2$ and in path 1 with probability $|\alpha_1|^2$. Measuring a superposition[6] $|\phi\rangle = \alpha_0 |0\rangle + \alpha_1 |1\rangle + \cdots + \alpha_j |j\rangle + \cdots + \alpha_n |n-1\rangle$ collapses it into an observed state $|j\rangle$, where the probability of observing state $|j\rangle$ is $|\alpha_j|^2$. Thus[7]

---

[6]Quantum mechanics requires that the superposition of states be a linear combination.
[7]Note that the absolute square of a complex number z, $|z|^2$, is defined to as follows: $|z|^2 = z\bar{z}$ where $\bar{z}$ denotes the *complex conjugate* of z and $|z|$ is the *complex modulus*.
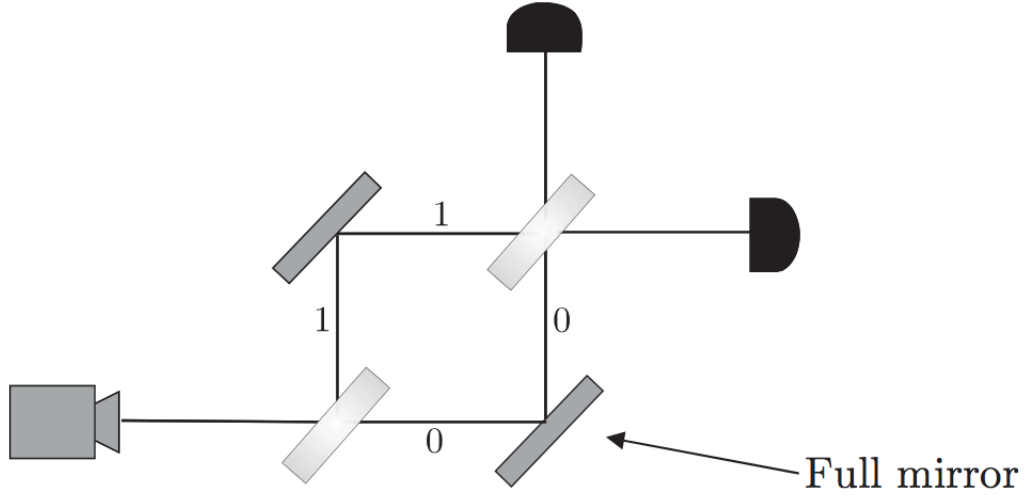
20

Figure 9: Experimental setup with two beam splitters: the Mach-Zehnder interferometer

$$\sum_{i=0}^{n-1} |\alpha_i|^2 = 1$$

That is, if we measure $|\phi\rangle$ we will observe the n-bit state state $|j\rangle$ with probability $|\alpha_j|^2$. In our example, $|\alpha_0|^2 + |\alpha_1|^2 = 1$ and

$$\alpha_0 \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \alpha_1 \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} \alpha_0 \\ \alpha_1 \end{pmatrix}$$

2.1 A Little More on Superposition

Consider some physical system that can be in N different, mutually exclusive classical states. Call these states $|0\rangle, |2\rangle, \cdots, |N-1\rangle$. A *classical state* is a state in which the system can be found if we observe it (with some details). A *pure quantum state* (usually just called state) $|\phi\rangle$ is a superposition of classical states, written[8]

$$|\phi\rangle = \alpha_0 |0\rangle + \alpha_1 |1\rangle + \cdots + \alpha_{N-1} |N-1\rangle$$

Here α_i is a complex number called the *amplitude* of $|i\rangle$. Intuitively, a system in quantum state $|\phi\rangle$ is in all classical states at the same time (!). In particular, it is in state $|0\rangle$

---

[8] $|i\rangle$ is the *ket* part of the Dirac *bra-ket* notation.

with amplitude α_0, in state $|1\rangle$ with amplitude α_1, and so on. Mathematically, the states $|0\rangle, \cdots, |N-1\rangle$ form an orthonormal basis of an N-dimensional Hilbert space (i.e., an N-dimensional vector space equipped with an inner product) of dimension N, and a quantum state $|\phi\rangle$ is a vector in this space.

2.2 Quantum Memory

In classical computation, the unit of information is a bit, which can be 0 or 1. In quantum computation, this unit is a quantum bit (*qubit*), which is a superposition of 0 and 1. Consider a system with 2 basis states, call them $|0\rangle$ and $|1\rangle$. We identify these basis states with the vectors $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$ respectively. As mentioned above, a single qubit can be in any superposition $|\phi\rangle = \alpha_0 |0\rangle + \alpha_1 |1\rangle$ with $|\alpha_0|^2 + |\alpha_1|^2 = 1$. This latter condition is sometimes called the *normalization constraint*. Again, $|\alpha_0|^2$ is the probability of finding the qubit in state $|0\rangle$. Similarly, $|\alpha_1|^2$ is the probability of finding the qubit in state $|1\rangle$.

Finally, it is worth noting that a complex amplitude α can be decomposed into the (unique) product $e^{i\theta}|\alpha|$, where $|\alpha|$ is the non-negative real number corresponding to the magnitude of α, and $e^{i\theta} = \frac{\alpha}{|\alpha|}$ has norm 1. θ is known as the phase and $e^{i\theta}$ is the phase factor. What is important about this is that the state vector described by $e^{i\theta}|\psi\rangle$ is equivalent to the state vector described by $|\psi\rangle$ where $e^{i\theta}$ is any complex vector with unit norm. For example, the state $|0\rangle + |1\rangle$ is equivalent to the state vector described by $e^{i\theta}|0\rangle + e^{i\theta}|1\rangle$.

A single qubit "lives" in the vector space $\mathbb{C}^2$. Systems of more than 1 qubit "live" in the tensor product space of several qubit systems. For example, a 2-qubit system has 4 basis states: $|0\rangle \otimes |0\rangle$, $|0\rangle \otimes |1\rangle$, $|1\rangle \otimes |0\rangle$, and $|1\rangle \otimes |1\rangle$. Here $|1\rangle \otimes |0\rangle$ means that the first qubit is in its basis state $|1\rangle$ and the second qubit is in its basis state $|0\rangle$.

More generally, a register of n qubits has 2^n basis states, each of the form

$$|b\rangle_0 \otimes |b\rangle_2 \otimes \cdots \otimes |b\rangle_{n-1}, \text{ with } b_i \in \{0, 1\} \tag{6}$$

We can abbreviate all of this as $|b_0 b_2 \cdots b_{n-1}\rangle$. In addition, since bit strings of length n can be viewed as numbers between 0 and $2^n - 1$, we can also write the basis states as $|0\rangle, |1\rangle, |2\rangle, \cdots |2^n - 1\rangle$. Thus a quantum register of n qubits can be in any superposition ϕ

$$|\phi\rangle = \alpha_0 |0\rangle + \alpha_1 |1\rangle + \cdots + \alpha_{2^n - 1} |2^n - 1\rangle \text{ where } \sum_{j=0}^{2^n - 1} |a_j|^2 = 1$$

3   Getting back to the experiment...

Lets call the beam splitter "gate" M_{BS}. Then

$$\begin{pmatrix} \alpha_0' \\ \alpha_1' \end{pmatrix} = M_{BS} \begin{pmatrix} \alpha_0 \\ \alpha_1 \end{pmatrix} = \begin{pmatrix} t_{11} & r_{21} \\ r_{12} & t_{22} \end{pmatrix} \begin{pmatrix} \alpha_0 \\ \alpha_1 \end{pmatrix}$$

where t_{ij} and r_{ij} are the amplitude transmission and reflection factors to and from the different ports of the beam splitter. Rearranging the complex coefficients t_{ij} and r_{ij} yields

$$M_{BS} = \frac{\exp(i\phi_0)}{\sqrt{2}} \begin{pmatrix} 1 & i \\ i & 1 \end{pmatrix}$$

where $i = \sqrt{-1}$ and ϕ_0 is an arbitrary phase constant usually set to zero.

Hence both of the transmitted light beams undergo a phase-shift of $\frac{\pi}{2}$ while the reflected beams are unaffected. In summary, when the photon passes through the beam splitter, we multiply its state vector by the splitter matrix

$$\frac{1}{\sqrt{2}} \begin{pmatrix} 1 & i \\ i & 1 \end{pmatrix}$$

Since we stipulate that the photon starts out in path 0, after passing through the first beam splitter it comes out in state

$$\frac{1}{\sqrt{2}} \begin{pmatrix} 1 & i \\ i & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ i \end{pmatrix}$$
$$= \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \frac{i}{\sqrt{2}} \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

This result corresponds with the observed behavior that after going through the first beam splitter, we would measure the photon in path 0 with probability $|\frac{1}{\sqrt{2}}|^2 = \frac{1}{2}$ and in path 1 with probability $|\frac{i}{\sqrt{2}}|^2 = \frac{1}{2}$.

However, if we do not measure which path the photon is in immediately after it passes through the first beam splitter, then its state remains[9]

---

[9]Note that $\begin{pmatrix} \alpha_0 \\ \alpha_1 \end{pmatrix}$ is alternate notation for $\begin{bmatrix} \alpha_0 \\ \alpha_1 \end{bmatrix}$.
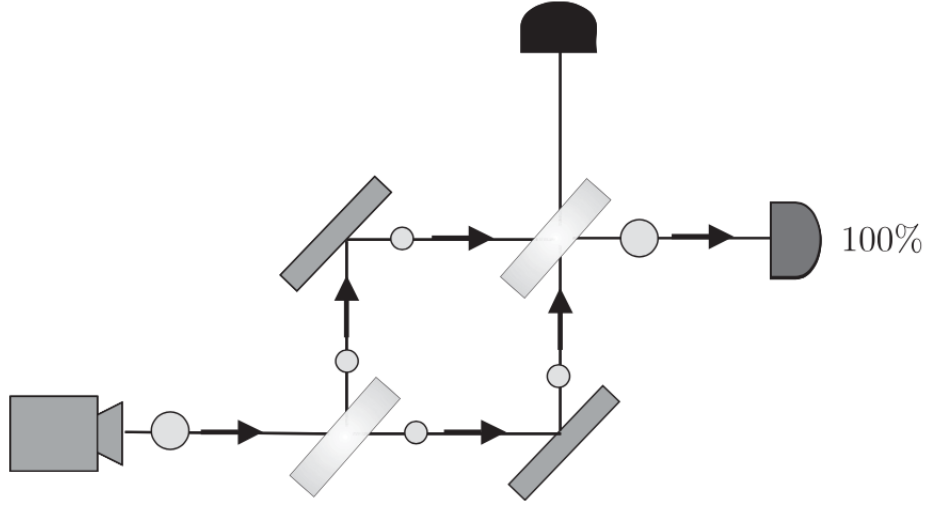
Figure 10: Experimental setup with two beam splitters

$$\frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ i \end{pmatrix}$$

Now (finally getting to the result), if the photon is allowed to pass through the second beam splitter before making any measurement of the photon's path its new state vector is

$$\left(\frac{1}{\sqrt{2}} \begin{pmatrix} 1 & i \\ i & 1 \end{pmatrix} \right) \left(\frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ i \end{pmatrix} \right) = \begin{pmatrix} 0 \\ i \end{pmatrix}$$

If we then measure the path of the photon **after** the second beam splitter (e.g. by the detectors shown in Figure 9), we find the photon on 1 path with probability $|i|^2 = 1$ and path 0 with probability $|0|^2 = 0$ (recall that $|\alpha_0|^2 + \alpha_1|^2 = 1$). Thus we expect that when we measure the photon after the second beam splitter it will be entirely on path 1, which is what we observe in experiments like the one illustrated in Figure 10.

Finally, the quantum mechanical interpretation of all of this is that the second beam splitter has caused the two paths (that were in *superposition*) to *interfere*, resulting in cancelation of the 0 path.

4 Acknowledgements

References

[1] Dorit Aharonov. Quantum computation.

[2] F. Gieres. Mathematical surprises and Dirac's formalism in quantum mechanics. *Reports on Progress in Physics*, 63:1893–1931, December 2000.

[3] Francois Henault. Quantum physics and the beam splitter mystery. https://arxiv.org/abs/1509.00393, September 2015.

[4] Wikipedia Contributors. Kronecker delta — Wikipedia, the free encyclopedia. `https://en.wikipedia.org/w/index.php?title=Kronecker_delta&oldid=862709627`, 2018. [Online; accessed 11-December-2018].

[5] K P Zetie, S F Adams, and R M Tocknell. How does a mach-zehnder interferometer work? *Physics Education*, 35(1):46, 2000.