Quantum Information Theory, 2024-5
Course Notes

Noah Linden & Paul Skrzypczyk

Contents

1 Mathematical Structures for Quantum Mechanics

1.1 Vectors, Dirac notation

States of a quantum system are vectors. We will use Dirac notation for these vectors and denote them as "kets":

$$|v\rangle \tag{1.1}$$

We start with a set of "basis" vectors

$$\{|0\rangle\ |1\rangle\ \ldots |n-1\rangle\} \tag{1.2}$$

and form complex linear combinations

$$|v\rangle = \sum_{j=0}^{n-1} v_j |j\rangle \qquad v_j \in \mathbb{C}. \tag{1.3}$$

We denote the set of such vectors $\mathbb{C}^n$. The complex numbers v_j are the components of $|v\rangle$ with respect to the given basis.

We can write these components as a column vector

$$\begin{bmatrix} v_0 \\ v_1 \\ \vdots \\ v_{n-1} \end{bmatrix}. \tag{1.4}$$

e.g. $\mathbb{C}^2$ is the set of vectors of the form

$$|v\rangle = \alpha|0\rangle + \beta|1\rangle \tag{1.5}$$

A system whose space of states is $\mathbb{C}^2$ is called a qubit.

We note that in order to specify the state of a qubit we potentially need an infinite number of bits: we will soon see that these qubit states will be normalized: $|\alpha|^2 + |\beta|^2 = 1$, so that for example $-1 \leq \mathrm{Re}(\alpha) \leq 1$. So consider representing $\mathrm{Re}(\alpha)$ in binary: we need potentially infinitely many digits after the binary point to specify $\mathrm{Re}(\alpha)$

1.2 Linear operators

Handout A Linear Operators

- A **linear operator** (on any vector space) takes vectors to vectors in such a way that

$$A(a|v\rangle + b|w\rangle) = a\,A|v\rangle + b\,A|w\rangle.$$

Thus one can define the action of an operator by its action on a basis.

- **Example** We can define the operator X on $\mathbb{C}^2$ by

$$X|0\rangle = |1\rangle; \qquad X|1\rangle = |0\rangle.$$

Thus the action of the operator X on any vector $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$ is given by

$$X|\psi\rangle = \alpha X|0\rangle + \beta X|1\rangle = \alpha|1\rangle + \beta|0\rangle.$$

This is the quantum analogue of the *NOT*.

• Suppose $|v_i\rangle$, $i = 0 \ldots n-1$ is a basis for $\mathbb{C}^n$; then we can define **the matrix of the operator A** with respect to the basis by

$$A|v_j\rangle = \sum_i A_{ij}|v_i\rangle.$$

• **Example** The matrix of X as defined above with respect to the basis

$$|v_1\rangle = |0\rangle = \begin{bmatrix} 1 \\ 0 \end{bmatrix}; \quad |v_2\rangle = |1\rangle = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

is

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

since

$$X|0\rangle = 0.|0\rangle + 1.|1\rangle; \quad X|1\rangle = 1.|0\rangle + 0.|1\rangle.$$

• Similarly given a matrix and a basis, we can define an operator associated to this matrix. For example given the matrix

$$\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix},$$

and using the standard basis

$$|0\rangle = \begin{bmatrix} 1 \\ 0 \end{bmatrix}; \quad |1\rangle = \begin{bmatrix} 0 \\ 1 \end{bmatrix},$$

we can define an operator Z on $\mathbb{C}^2$ by

$$Z|0\rangle = 1|0\rangle + 0.|1\rangle; Z|1\rangle = 0.|0\rangle - 1.|1\rangle.$$

• We will almost always use the standard basis for $\mathbb{C}^2$, and so will tend to think of operators and their matrices with respect to this basis interchangeably.

• An important set of operators are the **Pauli operators** X, Y, Z defined by

$$\begin{aligned} X|0\rangle &= |1\rangle; & X|1\rangle &= |0\rangle \\ Y|0\rangle &= i|1\rangle; & Y|1\rangle &= -i|0\rangle \\ Z|0\rangle &= |0\rangle; & Z|1\rangle &= -|1\rangle. \end{aligned}$$

Their associated matrices with respect to the standard basis

$$|v_1\rangle = |0\rangle; \quad |v_2\rangle = |1\rangle,$$

are

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}; \quad \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}; \quad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

• Two further operators on $\mathbb{C}^2$ which will be referred to later are the identity operator I

$$I|0\rangle = |0\rangle; \qquad I|1\rangle = |1\rangle,$$

and the operator, which depends on a real parameter θ and which I will denote $R(\theta)$, defined by

$$R(\theta)|0\rangle = \cos\theta|0\rangle + \sin\theta|1\rangle; \qquad R(\theta)|1\rangle = -\sin\theta|0\rangle + \cos\theta|1\rangle.$$

1.3 Inner product/Orthonormal basis

This is the analogue of the scalar product **a**.**b** of two vectors in three real dimensions.
Given two vectors $|v\rangle = \sum_{j=0}^{n-1} v_j|j\rangle$ and $|w\rangle = \sum_{j=0}^{n-1} w_j|j\rangle$, we denote the inner product of these two vectors as the Dirac bracket (bra-ket)

$$\langle w|v\rangle := \sum_{j=0}^{n-1} \overline{w}_j v_j \tag{1.6}$$

where $\overline{w}_j$ denotes the complex conjugate of w_j, and associated to the ket $|v\rangle = \sum_{j=0}^{n-1} v_j|j\rangle$, we introduce the "bra" vector

$$\langle v| = \sum_{j=0}^{n-1} \overline{v}_j\langle j|. \tag{1.7}$$

Just as the components of the ket vector can be written as a column vector - see equation (1.4) - the components of the bra vector correspond to the row vector

$$[\overline{v}_0, \overline{v}_1, \ldots, \overline{v}_{n-1}] \tag{1.8}$$

e.g. in $\mathbb{C}^2$,

$$|v\rangle = \alpha|0\rangle + \beta|1\rangle \quad \mapsto \quad \langle v| = \overline{\alpha}\langle 0| + \overline{\beta}\langle 1| \tag{1.9}$$

and given a second ket $|w\rangle = \gamma|0\rangle + \delta|1\rangle$, we form

$$\langle v|w\rangle = \Big(\overline{\alpha}\langle 0| + \overline{\beta}\langle 1|\Big)\Big(\gamma|0\rangle + \delta|1\rangle\Big) = \overline{\alpha}\gamma + \overline{\beta}\delta. \tag{1.10}$$

In $\mathbb{C}^2$ we *define* the inner products of the basis vectors as follows:

$$\langle 0|0\rangle = \langle 1|1\rangle = 1; \quad \langle 0|1\rangle = 0. \tag{1.11}$$

General properties of the inner product are covered in the following Handout.

Handout B Inner Product/ Orthonormal Bases

• An **inner product** takes two vectors (from the same vector space) and gives a complex number. We denote the inner product of $|u\rangle$ and $|v\rangle$ as

$$\langle u|v\rangle.$$

An inner product satisfies the following conditions

1. $\langle v|\Big[\alpha|w_1\rangle + \beta|w_2\rangle\Big] = \alpha\langle v|w_1\rangle + \beta\langle v|w_2\rangle$. We say the inner product is linear on the second factor
2. $\langle v|w\rangle = \overline{\langle w|v\rangle}$, where $\bar{}$ denotes the complex conjugate
3. $\langle v|v\rangle \geq 0$ with equality if and only if $|v\rangle$ is the zero vector (i.e. $\begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}$).

• A vector space with an inner product satisfying the above conditions is called a **Hilbert Space**.

• The above properties of the inner product mean that if $|z\rangle = \alpha|w_1\rangle + \beta|w_2\rangle$ then

$$\langle z|v\rangle = \overline{\alpha}\langle w_1|v\rangle + \overline{\beta}\langle w_2|v\rangle.$$

We say the inner product is conjugate-linear on the first factor.

Proof

$$\begin{aligned}
\langle z|v\rangle &= \overline{\langle v|z\rangle} \\
&= \overline{\langle v|\Big(\alpha|w_1\rangle + \beta|w_2\rangle\Big)} \\
&= \overline{\alpha\langle v|w_1\rangle + \beta\langle v|w_2\rangle} \\
&= \overline{\alpha}\overline{\langle v|w_1\rangle} + \overline{\beta}\overline{\langle v|w_2\rangle} \\
&= \overline{\alpha}\langle w_1|v\rangle + \overline{\beta}\langle w_2|v\rangle
\end{aligned}$$

• Two vectors $|u\rangle$ and $|v\rangle$ are said to be **orthogonal** if $\langle u|v\rangle = 0$

• The **norm** of a vector $|v\rangle$ is defined as $\| |v\rangle \| = (\langle v|v\rangle)^{\frac{1}{2}}$. A vector is normalised if $\| |v\rangle \| = 1$. Note that the vector $\frac{|v\rangle}{\||v\rangle\|}$ is normalised for any vector $|v\rangle$ (except the zero vector).

• A set of vectors $|v_i\rangle$ is **orthonormal** if

$$\begin{aligned} \langle v_i|v_j\rangle &= 1 \qquad i = j \\ &= 0 \qquad \text{otherwise.} \end{aligned}$$

• Thus if we express the vectors $|v\rangle$ and $|w\rangle$ with respect to an orthonormal basis $|i\rangle$ as

$$|v\rangle = \sum_i v_i|i\rangle, \quad |w\rangle = \sum_i w_i|i\rangle,$$

then the inner product is

$$\begin{aligned} \langle v|w\rangle &= \Big(\sum_i v_i|i\rangle, \sum_j w_j|j\rangle\Big) \\ &= \sum_{ij} \overline{v_i} w_j \langle i|j\rangle \\ &= \sum_i \overline{v_i} w_i. \end{aligned}$$

1.4 Dirac notation for operators

Handout C Dirac Notation for Operators

• We define

$$B = |u\rangle\langle v|$$

as a operator by giving its action on any vector $|w\rangle$ as

$$B|w\rangle = \Big[|u\rangle\langle v|\Big]|w\rangle = |u\rangle\langle v|w\rangle.$$

i.e. the action of B on $|w\rangle$ gives the vector $|u\rangle$ times the complex number $\langle v|w\rangle$.

• **Example** The operator $|0\rangle\langle 1|$ on $\mathbb{C}^2$ acts on the standard basis as

$$\begin{aligned} \Big[|0\rangle\langle 1|\Big]|0\rangle &= |0\rangle\langle 1|0\rangle \\ &= 0 \\ \Big[|0\rangle\langle 1|\Big])|1\rangle &= |0\rangle\langle 1|1\rangle \\ &= |0\rangle, \end{aligned}$$

and hence the action of the operator on an arbitrary vector in $\mathbb{C}^2$ is

$$\Big[|0\rangle\langle 1|\Big]\ \Big[\alpha|0\rangle + \beta|1\rangle\Big] = \beta|0\rangle.$$

Thus we may write the operator X as

$$X = |0\rangle\langle 1| + |1\rangle\langle 0|$$

• Consider an ON-basis $|i\rangle$ for $\mathbb{C}^n$. Recall that the matrix of an operator A with respect to this basis is defined by

$$A|k\rangle = \sum_{i=0}^{n-1} A_{ik}|i\rangle.$$

We can write the operator A in this basis as

$$A = \sum_{i,j=0}^{n-1} A_{ij}|i\rangle\langle j|.$$

This may easily be checked by considering the action of A on an arbitrary vector $|k\rangle$ in the basis:

$$\begin{aligned} A|k\rangle &= \Big[\sum_{i,j=0}^{n-1} A_{ij}|i\rangle\langle j|\Big]|k\rangle \\ &= \sum_{i,j=0}^{n-1} A_{ij}|i\rangle\langle j|k\rangle \\ &= \sum_{i=0}^{n-1} A_{ik}|i\rangle. \end{aligned}$$

• **Example** On $\mathbb{C}^2$, the operator A with the following matrix with respect to the standard orthonormal basis $|0\rangle$, $|1\rangle$:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

is

$$a|0\rangle\langle 0| + b|0\rangle\langle 1| + c|1\rangle\langle 0| + d|1\rangle\langle 1|$$

• Note also that the matrix elements of an operator A with respect to an ON-basis have a simple expression. For take the inner product of the equation $A|k\rangle = \sum_{i=0}^{n-1} A_{ik}|i\rangle$, with an element $|j\rangle$ of the ON-basis:

$$\langle j|A|k\rangle = \sum_{i=0}^{n-1} A_{ik}\langle j|i\rangle = A_{jk}.$$

1.5 Eigenvectors and eigenvalues/diagonal representation of an operator

Handout D Eigenvalues/Eigenvectors

• An **eigenvector** of an operator A is a vector $|v\rangle$ satisfying

$$A|v\rangle = \lambda|v\rangle,$$

for some complex number λ, the **eigenvalue** associated with that eigenvector. The eigenvalues and eigenvectors are most easily found by finding the matrix of A with respect to some ortho-normal basis, and finding the eigenvalues and eigenvectors of this matrix in the standard way (see the example below).

• A diagonal representation of an operator A on $\mathbb{C}^n$ is

$$A = \sum_{i=0}^{n-1} \lambda_i |v_i\rangle\langle v_i|,$$

where λ_i are eigenvalues of A and the vectors $|v_i\rangle$ form an orthonormal basis of the space. (All operators of interest to us in this course will have diagonal representations).

• Note that vectors $|v_i\rangle$ in the diagonal representation of A are eigenvectors of A, since consider one such vector $|v_k\rangle$:

$$\begin{aligned} A|v_k\rangle &= \left(\sum_{i=0}^{n-1} \lambda_i |v_i\rangle\langle v_i|\right)|v_k\rangle \\ &= \sum_{i=0}^{n-1} \lambda_i |v_i\rangle\langle v_i|v_k\rangle \\ &= \lambda_k |v_k\rangle. \end{aligned}$$

• **Example** Consider the operator X with matrix $\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ with respect to the standard basis. The equation for eigenvalues

$$\det(\sigma_x - \lambda I) = 0,$$

yields $\lambda = \pm 1$. The eigenvalue $\lambda = +1$ has normalised eigenvector $\frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ which corresponds to the ket vector $\frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$; let us call this vector $|+\rangle$. One may easily check that $X|+\rangle = |+\rangle$ as it should.

Similarly the eigenvalue $\lambda = -1$ has normalised eigenvector $\frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ -1 \end{pmatrix}$ which corresponds to the ket vector $\frac{1}{\sqrt{2}}(|0\rangle - |1\rangle)$; let us call this vector $|-\rangle$.

Thus the diagonal representation will be

$$\begin{aligned} X &= \sum \lambda_i |v_i\rangle\langle v_i| \\ &= |+\rangle\langle +| \quad - \quad |-\rangle\langle -| \\ &= \left(\frac{|0\rangle + |1\rangle}{\sqrt{2}}\right)\left(\frac{\langle 0| + \langle 1|}{\sqrt{2}}\right) - \left(\frac{|0\rangle - |1\rangle}{\sqrt{2}}\right)\left(\frac{\langle 0| - \langle 1|}{\sqrt{2}}\right). \end{aligned}$$

By multiplying the brackets out it may be checked that this expression is indeed equal to

$$|0\rangle\langle 1| + |1\rangle\langle 0|,$$

the expression for X we have seen before.

If an operator has the property that more than one orthogonal eigenstate correspond to the same eigenvalue, then this eigenvalue is said to be degenerate. e.g. consider the following operator on $\mathbb{C}^3$:

$$|0\rangle\langle 0| + |1\rangle\langle 1| - |2\rangle\langle 2|. \tag{1.12}$$

$\{|0\rangle, |1\rangle, |2\rangle\}$ is the standard orthonormal basis for $\mathbb{C}^3$.

The states $|0\rangle$ and $|1\rangle$ correspond to the same eigenvalue $+1$, and this eigenvalue is degenerate. We say the eigenspace corresponding to $\lambda = +1$ is 2 dimensional since it is spanned by two eigenvectors.

1.6 Adjoint of an operator/self-adjoint operators

Handout E Adjoints

• Given an operator A, we define its **adjoint**, $A^\dagger$ via

$$\langle v|A^\dagger|w\rangle = \langle w|A|v\rangle^*$$

for all vectors $|v\rangle$ and $|w\rangle$.

• In particular consider an ON-basis, $|i\rangle$, then

$$\langle i|A^\dagger|j\rangle = \langle j|A|i\rangle^*,$$

which defines a relationship between the components of the matrix of A with respect to that basis, and the components of $A^\dagger$.

• **Example** In $\mathbb{C}^2$, if A has matrix

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix},$$

then $A^\dagger$ has matrix

$$\begin{pmatrix} a^* & c^* \\ b^* & d^* \end{pmatrix},$$

• **Example** X has matrix

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

with respect to the standard basis, thus $X^\dagger$ has matrix

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

We say that X is **self-adjoint**, or **hermitian**. You may check that Y and Z are also self-adjoint, but that $R(\theta)^\dagger = R(-\theta)$.

• Recall that to a vector $|v\rangle$ we associated a dual vector $\langle v|$. We write

$$\Big(|v\rangle\Big)^\dagger = \langle v|.$$

We note that the dual vector to $A|v\rangle$ is $\langle v|A^\dagger$. To see this write

$$\langle u|w\rangle = \langle w|u\rangle^*,$$

now call $|u\rangle = A|v\rangle$, so

$$\langle u|w\rangle = \langle w|A|v\rangle^* = \langle v|A^\dagger|w\rangle.$$

Since this must be true for all vectors $|w\rangle$, we see that $\langle u| = \langle v|A^\dagger$.

It is a useful fact that

$$(|a\rangle\langle b|)^\dagger = |b\rangle\langle a|, \tag{1.13}$$

since let $A = |a\rangle\langle b|$, and consider two arbitrary states $|v\rangle$ and $|w\rangle$. Then [with $*$ denoting complex conjugate]

$$\begin{aligned} \langle v|A^\dagger|w\rangle &= \Big(\langle w|A|v\rangle\Big)^* \\ &= \Big(\langle w|a\rangle\langle b|v\rangle\Big)^* \\ &= \langle w|a\rangle^*\langle b|v\rangle^* \\ &= \langle a|w\rangle\langle v|b\rangle \\ &= \langle v|b\rangle\langle a|w\rangle \\ &= \langle v|\Big(|b\rangle\langle a|\Big)|w\rangle. \end{aligned} \tag{1.14}$$

Since this is true for all $|v\rangle$ and $|w\rangle$,

$$(|a\rangle\langle b|)^\dagger = |b\rangle\langle a|. \tag{1.15}$$

1.7 Unitary operators/Identity operator

Handout F Unitary Operators

• An operator U (on any vector space) is said to be **unitary** if it satisfies

$$UU^\dagger = U^\dagger U = 1$$

• **Example** X, Y, Z and I defined on $\mathbb{C}^2$ are unitary.

• The definition of unitarity means that unitary operators are invertible (i.e. their action can be "undone"), since if one follows the action of U by $U^\dagger$, any vector transforms as follows:

$$|v\rangle \mapsto U|v\rangle \mapsto U^\dagger U|v\rangle = |v\rangle.$$

Note that this condition of reversibility is quite different from classical gates.

• Unitary operations preserve the inner product between vectors for let us define

$$|v_U\rangle := U|v\rangle \quad \text{and} \quad |w_U\rangle := U|w\rangle$$

then

$$\begin{aligned} \langle v_U|w_U\rangle &= [U|v\rangle]^\dagger U|w\rangle \\ &= \langle v|U^\dagger U|w\rangle \\ &= \langle v|w\rangle. \end{aligned}$$

We note that for any orthonormal basis for $\mathbb{C}^n$,

$$\{|e_0\rangle, |e_2\rangle \ldots |e_{n-1}\rangle\} \tag{1.16}$$

the following is a representation of the identity operator:

$$\sum_{j=0}^{n-1} |e_j\rangle\langle e_j|. \tag{1.17}$$

Note that the identity is self-adjoint.
Thus for example on $\mathbb{C}^2$

$$|+\rangle\langle +| + |-\rangle\langle -| = |0\rangle\langle 0| + |1\rangle\langle 1| = I. \tag{1.18}$$

And the following is an example of a unitary operator on $\mathbb{C}^2$:

$$U = |+\rangle\langle 0| + |-\rangle\langle 1|. \tag{1.19}$$

It can be checked that

$$UU^\dagger = |+\rangle\langle +| + |-\rangle\langle -| = I. \tag{1.20}$$

We note that this example of a unitary U illustrates the general property of a unitary in any dimension: it takes an orthonormal basis to an orthonormal basis.

1.8 Projection operators

Handout G Projection Operators

• For any orthonormal set of vectors $|v_i\rangle$, $i = 0 \ldots k-1$ in $\mathbb{C}^n$, where k may be less than the dimension n, the operator

$$P = \sum_{i=0}^{k-1} |v_i\rangle\langle v_i|$$

is called a **projection operator** onto the space spanned by the vectors $|v_i\rangle$, $i = 0 \ldots k-1$.

• **Example** In $\mathbb{C}^2$, the operator $|0\rangle\langle 0|$ is a projector: its action on an arbitrary vector $|v\rangle = \alpha|0\rangle + \beta|1\rangle$, is

$$|0\rangle\langle 0|\Big[\alpha|0\rangle + \beta|1\rangle\Big] = \alpha|0\rangle$$

(i.e. it projects out the component of $|v\rangle$ along $|0\rangle$); hence the name projector.

• **Example** In $\mathbb{C}^2$, the operator $|+\rangle\langle +|$ is a projector; indeed for any vector $|v\rangle$, $|v\rangle\langle v|$ projects a vector along the direction $|v\rangle$.

• **Example** In $\mathbb{C}^3$, consider the orthonormal basis for the whole space: $|0\rangle$, $|1\rangle$, $|2\rangle$. Then for example

$$|0\rangle\langle 0| + |1\rangle\langle 1|$$

is a projector onto the two-dimensional subspace spanned by $|0\rangle$, $|1\rangle$. Note that

$$\Big[|0\rangle\langle 0| + |1\rangle\langle 1|\Big]\Big[\alpha|0\rangle + \beta|1\rangle + \gamma|2\rangle\Big] = \alpha|0\rangle + \beta|1\rangle$$

• You should check that projection operators, as defined above satisfy the two conditions

$$P = P^\dagger; \qquad P^2 = P.$$

In fact these two conditions are sometimes used as the definition of a projection operator, since any operator satisfying them may be written as

$$\sum_{i=0}^{k-1} |v_i\rangle\langle v_i|$$

for suitably chosen vectors $|v_i\rangle$.

• Finally note that $1 - P$ is also a projector if P is.

1.9 Spectral theorem for self-adjoint operators

Handout H Spectral Theorem for Self-adjoint Operators

• **Spectral Theorem**: Self-adjoint operators are diagonalizable and have real eigenvalues. Thus any self-adjoint operator A on $\mathbb{C}^n$ may be written as

$$A = \sum_{i=0}^{n-1} \lambda_i |v_i\rangle\langle v_i|$$

for some orthonormal basis of the space $|v_i\rangle$ $i = 0 \ldots n-1$. Equivalently we may group together the eigenvectors with same eigenvalues and write

$$A = \sum_{\lambda_\alpha \text{distinct}} \lambda_\alpha P_\alpha$$

where P_α denotes the projection operator onto eigenspace associated to the eigenvalue λ_α.

• We will not prove the spectral theorem but note that the second part is easy since if $|v\rangle$ is a normalised eigenvector of A with eigenvalue λ, then

$$A|v\rangle = \lambda|v\rangle \Rightarrow \langle v|A|v\rangle = \lambda\langle v|v\rangle = \lambda;$$

but

$$\Big[\langle v|A|v\rangle\Big]^* = \Big[\langle v|A^\dagger|v\rangle\Big] = \Big[\langle v|A|v\rangle\Big],$$

so λ is real.

• **Example** We showed earlier in the course that

$$X = |+\rangle\langle +| \quad - \quad |-\rangle\langle -|$$

• **Example**
The operator A on $\mathbb{C}^4$ given by

$$A = |0\rangle\langle 0| + |1\rangle\langle 1| - |2\rangle\langle 2| - |3\rangle\langle 3|$$

has two degenerate eigenvalues and may be written as

$$A = \Big[|0\rangle\langle 0| + |1\rangle\langle 1|\Big] - \Big[|2\rangle\langle 2| + |3\rangle\langle 3|\Big]$$

(i.e. we have written it as a sum of eigenvalues times projection operators)

2 The Rules of Quantum Mechanics

2.1 States; evolution; measurement

Handout I The rules of quantum mechanics

1. States of a system are described by normalised vectors in Hilbert space; two vectors differing by an overall phase are equivalent (i.e. $|\psi_1\rangle \equiv |\psi_2\rangle$ if $|\psi_1\rangle = e^{i\theta}|\psi_2\rangle$). One says that quantum states correspond to **rays** in Hilbert space.

2. Evolutions of a system are described by unitary operators.

3. (Von Neumann's projection postulate). Observables (measurements) correspond to self-adjoint operators. Furthermore, when a measurement corresponding to an operator A is made, the numerical outcome of the measurement is an eigenvalue of A (λ_α, say) and immediately after the measurement, the state of the system is associated eigenstate.

   Let us write A in its spectral decompostion:

   $$A = \sum_{\lambda_\alpha \text{distinct}} \lambda_\alpha P_\alpha$$

   If the state of the system before measurement is $|\psi\rangle$, then the probability that the eigenvalue λ_α is observed is

   $$\text{Prob}(\lambda_\alpha) = \| P_\alpha|\psi\rangle \|^2 = \langle\psi|P_\alpha P_\alpha|\psi\rangle = \langle\psi|P_\alpha|\psi\rangle,$$

   and the (normalised) state after the measurement is

   $$\frac{P_\alpha|\psi\rangle}{\left[\langle\psi|P_\alpha|\psi\rangle\right]^{1/2}}.$$

   The expected value of the observable A in the state ψ is

   $$\langle\psi|A|\psi\rangle = \sum_\alpha \lambda_\alpha \,\text{Prob}(\lambda_\alpha)$$

Example

Let the state before measurement be $|\psi\rangle = |0\rangle \in \mathbb{C}^2$, and let us measure the observable X.

As we have calculated earlier, the eigenvalues of X are $+1, -1$ with associated projectors $P_{+1} = |+\rangle\langle+|$ and $P_{-1} = |-\rangle\langle-|$, thus the spectral decompostion of X is

$$X = |+\rangle\langle+| \quad - \quad |-\rangle\langle-| = P_{+1} - P_{-1},$$

where

$$|+\rangle = \frac{|0\rangle + |1\rangle}{\sqrt{2}}; \quad |-\rangle = \frac{|0\rangle - |1\rangle}{\sqrt{2}}.$$

Thus, the probability that the measurement outcome is $+1$ is

$$\langle\psi|P_{+1}|\psi\rangle = \langle 0|\Big[|+\rangle\langle +|\Big]|0\rangle = \frac{1}{2},$$

and the state immediately after the measurement is

$$\frac{P_\alpha|\psi\rangle}{\Big[\langle\psi|P_\alpha|\psi\rangle\Big]^{1/2}} = |+\rangle.$$

• We note that rule 3 can be interpreted as follows. Consider a self-adjoint operator A on $\mathbb{C}^n$ and its eigenvectors $|v_i\rangle$, $i = 0 \ldots n-1$ (which form an orthonormal basis for $\mathbb{C}^n$). Let us write the state $|\psi\rangle$ using this basis:

$$|\psi\rangle = \sum_{i=0}^{n-1} a_i|v_i\rangle.$$

Consider now the eigenspace, of dimension k, say, associated to the eigenvalue λ_α; the associated projector is

$$P_\alpha = \sum_{i=0}^{k-1} |v_i\rangle\langle v_i|;$$

(we have labelled the basis elements $|v_i\rangle$ such P_α projects onto the first k eigenvectors).
The probability that the eigenvalue λ_α is found when A is measured is

$$\begin{aligned}
&\langle\psi|P_\alpha|\psi\rangle \\
&= \Big[\sum_{i=0}^{n-1} a_i^*\langle v_i|\Big]\Big[\sum_{j=0}^{k-1}|v_j\rangle\langle v_j|\Big]\Big[\sum_{m=0}^{n-1} a_m|v_m\rangle\Big] \\
&= \Big[\sum_{i=0}^{n-1} a_i^*\langle v_i|\Big]\Big[\sum_{m=0}^{k-1} a_m|v_m\rangle\Big] \\
&= \sum_{m=0}^{k-1}|a_m|^2.
\end{aligned}$$

The complex numbers a_i are called **probability amplitudes**.

• **Example** In the example above, we may write the state $|0\rangle$ in terms of the eigenstates of X as

$$|0\rangle = \frac{1}{\sqrt{2}}\Big[|+\rangle + |-\rangle\Big].$$

Thus the probability of finding the eigenvalue $+1$ when measuring X is $|\frac{1}{\sqrt{2}}|^2 = \frac{1}{2}$, as we found before.

When computing what happens when the outcome λ occurs when making a measurement of an operator A on the state $|\psi\rangle$, it is often helpful to compute

$$P|\psi\rangle \tag{2.1}$$

where P is the projection operator associated to λ in the spectral decomposition of the measurement operator. This is because the probability that this outcome occurred is $||P|\psi\rangle||^2$ and the state

after the measurement is the normalized version of the state:

$$\frac{P|\psi\rangle}{||P|\psi\rangle||}. \tag{2.2}$$

We note that measurement typically disturbs the state. And the state is unchanged if and only if the state $|\psi\rangle$ is an eigenstate of the measurement operator A. If the state is an eigenstate of A with eigenvalue λ_j then the value obtained on measurement of A is λ_j with certainty [i.e. probability 1].

e.g. consider the general state $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$ on $\mathbb{C}^2$. Let us measure $Z = |0\rangle\langle 0| - |1\rangle\langle 1|$. Then the probability of getting the eigenvalue $\lambda = +1$ is $|\alpha|^2$ and in this case the state after the measurement is $|0\rangle$; and the probability of getting the eigenvalue $\lambda = -1$ is $|\beta|^2$ and in this case the state after the measurement is $|1\rangle$. In the case that the state was $|0\rangle$ or $|1\rangle$ we get the associated eigenvalue ($+1$ and -1 respectively) with certainty and the state is unchanged by the measurement. For typical values of α and β the state changes after the measurement (i.e. the measurement disturbs the state).

2.2 Distinguishing orthogonal and non-orthogonal states

Consider that a source emits a qubit in either the state $|0\rangle$ or $|1\rangle$. Can the detector make a measurement to find out which state the source sent, with certainty?

This is easy. For example the detector could measure $Z = |0\rangle\langle 0| - |1\rangle\langle 1|$. Or indeed any measurement operator of the form

$$A = \lambda_1|0\rangle\langle 0| + \lambda_2|1\rangle\langle 1| \quad \lambda_1 \neq \lambda_2, \tag{2.3}$$

since if $|\psi\rangle = |0\rangle$, for example, $\text{Prob}(\lambda_1) = ||\Big(|0\rangle\langle 0|\Big)|\psi\rangle||^2 = 1$ and $\text{Prob}(\lambda_2) = ||\Big(|1\rangle\langle 1|\Big)|\psi\rangle||^2 = 0$.

We have effectively sent a bit using a qubit.

More generally, consider that the source sends one of two general orthogonal and normalized states $|u\rangle$ and $|v\rangle$, i.e. $\langle u|v\rangle = 0$. Then the detector can measure

$$A = \lambda_1|u\rangle\langle u| + \lambda_2|v\rangle\langle v| \quad \lambda_1 \neq \lambda_2, \tag{2.4}$$

to determine which state was sent, with certainty.

Now we consider what happens if the source again sends one of two states, but they are not orthogonal. e.g. consider that the source sends a qubit in either the state $|0\rangle$ or the state $|+\rangle$. Can the detector distinguish them perfectly? Consider that the detector measures Z.

If the state were $|0\rangle$ then $\text{Prob}(\lambda = +1) = 1$ and $\text{Prob}(\lambda = -1) = 0$. But if the detector meaures Z and gets $\lambda = +1$ they cannot be sure that the state was $|0\rangle$, since if the state were $|+\rangle$ the outcome $\lambda = +1$ can also occur.

But if the outcome $\lambda = -1$ occurs, then the detector can be certain about which state was sent since only $|+\rangle$ can give this outcome. So if $\lambda = -1$ occurs, the state cannot have been $|0\rangle$. [Although notice that half the time that the source sends $|+\rangle$ the measurement of Z yields $\lambda = +1$, so the detector cannot be certain which state was sent.]

2.3 Interference

Consider the state $|0\rangle$ on $\mathbb{C}^2$. If we measure X we find $\text{Prob}(\lambda = -1) = 1/2$. Similarly if the state is $|1\rangle$ and we measure X we also find that $\text{Prob}(\lambda = -1) = 1/2$. Now consider the following superposition of these two states:

$$|\psi\rangle = \frac{|0\rangle + |1\rangle}{\sqrt{2}} = |+\rangle. \tag{2.5}$$

Now if we measure X we find $\text{Prob}(\lambda = -1) = 0$. Thus if we have two states and superpose them, the probabilities of the outcomes of a given measurement do not add. In this example we have "destructive" interference - a wave-like property of quantum mechanics.

In genearl the probability of an event depends on the *relative* phase between the terms $|0\rangle$ and $|1\rangle$. Thus the relative phase [unlike the overall phase] has physical relevance.

For example if the state we start with is

$$|\psi\rangle = \frac{|0\rangle + e^{i\phi}|1\rangle}{\sqrt{2}}, \tag{2.6}$$

then we find $\text{Prob}(\lambda = -1) = \frac{1}{2}(1 - \cos\phi)$ when we measure X.

3 Multi-party quantum systems

3.1 The Hilbert space of two-party quantum systems - the tensor product

Handout J Two party quantum systems

• We aim initially to describe the space of states of two systems each of which has a qubit as its space of states. This is the quantum version of the space of two bits. Amongst these states are states in which the first system is in state $|\psi_A\rangle$ and the second in $|\psi_B\rangle$ respectively. We will soon see that the linearity of quantum mechanics means that that not all states of the combined system are of this form.

• We introduce the **tensor product**, denoted $\mathbb{C}^2 \otimes \mathbb{C}^2$ of the spaces of each of the qubits. One way of describing the space is to say it is the vector space spanned by the four basis vectors

$$|0\rangle \otimes |0\rangle$$
$$|0\rangle \otimes |1\rangle$$
$$|1\rangle \otimes |0\rangle$$
$$|1\rangle \otimes |1\rangle.$$

The vector $|0\rangle \otimes |0\rangle$ is often written $|0\rangle|0\rangle$ or just $|00\rangle$ and so on.

• Thus an arbitrary vector in the space is a linear combination of these basis vectors:

$$|\psi\rangle = a_{00}|00\rangle + a_{01}|01\rangle + a_{10}|10\rangle + a_{11}|11\rangle = \sum_{i,j=0}^{1} a_{ij}|ij\rangle.$$

The vectors $|00\rangle, |01\rangle, |10\rangle, |11\rangle$ correspond to

$$\begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}; \quad \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}; \quad \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}; \quad \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

respectively.

• If $|\psi_A\rangle = a_0|0\rangle + a_1|1\rangle$ describes the state of the first system and $|\psi_B\rangle = b_0|0\rangle + b_1|1\rangle$ the state of the second then the state of the combined system is

$$\begin{aligned} |\psi_A\rangle \otimes |\psi_B\rangle &= \Big[a_0|0\rangle + a_1|1\rangle\Big] \otimes \Big[b_0|0\rangle + b_1|1\rangle\Big] \\ &= a_0b_0|00\rangle + a_0b_1|01\rangle + a_1b_0|10\rangle + a_1b_1|11\rangle. \end{aligned}$$

The following natural rule (part of the formal definition of the tensor product) was used

$$\begin{aligned} &\Big[|v_1\rangle + |v_2\rangle\Big] \otimes \Big[|w_1\rangle + |w_2\rangle\Big] \\ &= |v_1\rangle|w_1\rangle + |v_1\rangle|w_2\rangle + |v_2\rangle|w_1\rangle + |v_2\rangle|w_2\rangle. \end{aligned}$$

Thus we may write

$$\begin{bmatrix} a_0 \\ a_1 \end{bmatrix} \otimes \begin{bmatrix} b_0 \\ b_1 \end{bmatrix} = \begin{bmatrix} a_0 b_0 \\ a_0 b_1 \\ a_1 b_0 \\ a_1 b_1 \end{bmatrix}$$

We say that states of the form $|\psi_A\rangle \otimes |\psi_B\rangle$ are of **product form** or **direct product** states.

• The inner product of two states in the tensor product is defined as follows: first we define the inner product on states of product form:

$$\Big(|v_1\rangle|v_2\rangle, |w_1\rangle|w_2\rangle\Big) = \langle v_1|w_1\rangle\langle v_2|w_2\rangle,$$

then use the fact that an inner product is linear on the second factor (conjugate linear on the first), thus

$$\Big(|v_1\rangle|v_2\rangle, \Big[|w_1\rangle|w_2\rangle + |z_1\rangle|z_2\rangle\Big]\Big) = \langle v_1|w_1\rangle\langle v_2|w_2\rangle + \langle v_1|z_1\rangle\langle v_2|z_2\rangle.$$

Particular cases are

$$\Big(|0\rangle|0\rangle, |0\rangle|0\rangle\Big) = 1; \quad \Big(|0\rangle|0\rangle, |0\rangle|1\rangle\Big) = 0; \quad \text{etc.}$$

• It is fundamentally important that not all quantum states of two (or more) systems are of product form.
e.g.

$$\begin{aligned} |\Psi_1\rangle &= \frac{|00\rangle + |11\rangle}{\sqrt{2}} \\ \text{or} \quad |\Psi_2\rangle &= \frac{|01\rangle - |10\rangle}{\sqrt{2}}. \end{aligned}$$

Since consider $|\Psi_1\rangle$; if it were of product form then it could be written as

$$\begin{aligned} &\Big[a_0|0\rangle + a_1|1\rangle\Big] \otimes \Big[b_0|0\rangle + b_1|1\rangle\Big], \\ &= a_0 b_0|0\rangle|0\rangle + a_0 b_1|0\rangle|1\rangle + a_1 b_0|1\rangle|0\rangle + a_1 b_1|1\rangle|1\rangle \end{aligned}$$

for some choice of the parameters a_0, a_1, b_0, b_1. However looking at the $|0\rangle|1\rangle$ term we see that a_0 or b_1 must be zero and then either the $|0\rangle|0\rangle$ or the $|1\rangle|1\rangle$ term would be zero.
Any state which is not of product form is said to be **entangled**.

• The tensor product of more general Hilbert spaces is defined as follows: e.g. consider the tensor product of $\mathbb{C}^n$ with $\mathbb{C}^m$. We may define the tensor product as the space spanned by

$$|v_i\rangle \otimes |w_j\rangle$$

where $|v_i\rangle$, $i = 0 \ldots n-1$ is a basis for $\mathbb{C}^n$, and $|w_j\rangle$, $j = 0 \ldots m-1$ is a basis for $\mathbb{C}^m$. The dimension of the vector space is mn.

3.2 Multi-party quantum systems

The Hilbert space of n qubits is

$$\mathcal{H} = \mathbb{C}^2 \otimes \mathbb{C}^2 \otimes \ldots \otimes \mathbb{C}^2 = \mathbb{C}^{2^n} \tag{3.1}$$

It is spanned by the 2^n states $|0\rangle \otimes |0\rangle \ldots |0\rangle$ to $|1\rangle \otimes |1\rangle \ldots |1\rangle$, and a general state is a complex superposition of these states.

3.3 Operators on tensor product states

If operator A acts on the Hilbert space of the first system and operator B acts on the Hilbert space of the second system, then we define the action of the operator $A \otimes B$ on the tensor product space as follows: let $|\psi\rangle$ be a general state on the tensor product space:

$$|\psi\rangle = \sum_{i,j} \alpha_{ij} |v_i\rangle |w_j\rangle \tag{3.2}$$

where $\{|v_i\rangle\}$ is a basis for the first Hilbert space and $\{|w_j\rangle\}$ is a basis for the second Hilbert space. Then the operator $A \otimes B$ is defined by

$$A \otimes B|\psi\rangle = \sum_{i,j} \alpha_{ij} A|v_i\rangle \otimes B|w_j\rangle \tag{3.3}$$

(i.e. $A \otimes B$ acts linearly). Note that where there is no confusion, we can write

$$A|v_i\rangle \otimes B|w_j\rangle \quad \text{as} \quad A|v_i\rangle B|w_j\rangle \quad \text{etc.} \tag{3.4}$$

e.g. $X \otimes Z$ on $\mathbb{C}^2 \otimes \mathbb{C}^2$:

$$\begin{aligned} X \otimes Z|0\rangle \otimes |0\rangle &= X|0\rangle \otimes Z|0\rangle \equiv X|0\rangle Z|0\rangle = |1\rangle|0\rangle; \\ X \otimes Z|1\rangle|1\rangle &= X|1\rangle Z|1\rangle = -|0\rangle|1\rangle. \end{aligned} \tag{3.5}$$

Thus for example

$$X \otimes Z \frac{|0\rangle|0\rangle + |1\rangle|1\rangle}{\sqrt{2}} = \frac{|1\rangle|0\rangle - |0\rangle|1\rangle}{\sqrt{2}}. \tag{3.6}$$

An important example is where Alice acts with X, say, and Bob does nothing. This corresponds to the operator $X \otimes I$ on the whole Hilbert space.

The most general operator on a tensor product space is a sum of operators of the form $A \otimes B$, e.g. $X \otimes Z + Z \otimes X$ on $\mathbb{C}^2 \otimes \mathbb{C}^2$. e.g.

$$\Big(X \otimes Z + Z \otimes X\Big)|0\rangle|0\rangle = |1\rangle|0\rangle + |0\rangle|1\rangle. \tag{3.7}$$

Any operator that can be written in the form $A \otimes B$ is called a local operator; it corresponds to Alice and Bob each acting in their own Hilbert space, without interaction.

Note that a local operator takes product states to product states: for any $|\phi\rangle$ and $|\eta\rangle$:

$$A \otimes B|\phi\rangle|\eta\rangle = A|\phi\rangle \otimes B|\eta\rangle. \tag{3.8}$$

But an operator that takes a particular orthonormal basis of product states to an orthonormal basis of product states is not necessarily a local operator. A famous example is the controlled-not $CNOT$ on $\mathbb{C}^2 \otimes \mathbb{C}^2$:

$$\begin{aligned} CNOT|0\rangle|0\rangle &= |0\rangle|0\rangle \\ CNOT|0\rangle|1\rangle &= |0\rangle|1\rangle \\ CNOT|1\rangle|0\rangle &= |1\rangle|1\rangle \\ CNOT|1\rangle|1\rangle &= |1\rangle|0\rangle. \end{aligned} \tag{3.9}$$

Equivalence under local unitary transformations.
Two states $|\Psi_1\rangle$ and $|\Psi_2\rangle$ are said to be equivalent under local unitary transformations if

$$|\Psi_2\rangle = U_1 \otimes U_2 |\Psi_1\rangle. \tag{3.10}$$

As will become clear such states have equivalent amounts of entanglement.
Thus the Bell states are equivalent under local unitary transformations. e.g.

$$\frac{|0\rangle|1\rangle - |1\rangle|0\rangle}{\sqrt{2}} = I \otimes X \frac{|0\rangle|0\rangle - |1\rangle|1\rangle}{\sqrt{2}}. \tag{3.11}$$

3.4 The EPR state

We consider the scenario where a source sends one qubit to Alice and a second to Bob. The pair of particles is in the Einstein-Podolsky-Rosen (EPR) state:

$$|\psi\rangle = \frac{|0\rangle|1\rangle - |1\rangle|0\rangle}{\sqrt{2}}. \tag{3.12}$$

Let us consider that Alice measures Z on her qubit. This corresponds to $Z \otimes I$ on the whole Hilbert space. In order to understand the outcomes of the measurement, we need to write this operator in spectral decomposition:

$$Z \otimes I = \Big(|0\rangle\langle 0| - |1\rangle\langle 1|\Big) \otimes I = |0\rangle\langle 0| \otimes I - |1\rangle\langle 1| \otimes I := P_+ - P_-. \tag{3.13}$$

Let us consider the case that the eigenvalue is $\lambda = +1$:

$$P_+ |\psi\rangle = |0\rangle\langle 0| \otimes I |\psi\rangle = \frac{1}{\sqrt{2}} |0\rangle|1\rangle. \tag{3.14}$$

Thus the probability of this outcome is $\frac{1}{2}$ and state after the measurement, if this outcome occurred, is

$$|0\rangle|1\rangle. \tag{3.15}$$

Consider that now Bob also measures Z on his qubit, with the pair of particles now in the post-measurement state $|0\rangle|1\rangle$. This corresponds to $I \otimes Z$ on the full Hilbert space:

$$I \otimes Z = I \otimes \Big(|0\rangle\langle 0| - |1\rangle\langle 1|\Big) = I \otimes |0\rangle\langle 0| - I \otimes |1\rangle\langle 1| := Q_+ - Q_-. \tag{3.16}$$

Let us consider the situation that Bob gets $\lambda = -1$:

$$Q_- |0\rangle|1\rangle = I \otimes |1\rangle\langle 1| = |0\rangle|1\rangle. \tag{3.17}$$

Thus the probability of this outcome is 1. i.e. if Alice gets $\lambda = +1$ (which occurs with probability half), then Bob gets $\lambda = -1$ *with certainty*. It is not difficult to show, that if Alice had got $\lambda = -1$ then Bob would have got $\lambda = +1$ with certainty. Thus we have exact anti-correlation. indeed this exact anti-correlation occurs if Alice and Bob both measure the same operator with eigenvalues +1 and -1 (examples include Alice and Bob both measuring X or both measuring Y).

We now consider that Alice and Bob, rather than having particles in the singlet state, each have a qubit with the qubits being in the general state

$$|\Psi\rangle = \alpha_{00}|0\rangle|0\rangle + \alpha_{01}|0\rangle|1\rangle + \alpha_{10}|1\rangle|0\rangle + \alpha_{11}|1\rangle|1\rangle. \tag{3.18}$$

We consider again that Alice measures Z. It is convenient to write $|\Psi\rangle$ as

$$|\Psi\rangle = |0\rangle(\alpha_{00}|0\rangle + \alpha_{01}|1\rangle) + |1\rangle(\alpha_{10}|0\rangle + \alpha_{11}|1\rangle). \tag{3.19}$$

We want to find out what happens if Alice gets the eigenvalue $\lambda = +1$. Then writing $Z \otimes I = P_+ - P_-$, as before, we can calculate that

$$P_+|\Psi\rangle = |0\rangle\langle 0| \otimes I|\Psi\rangle = |0\rangle(\alpha_{00}|0\rangle + \alpha_{01}|1\rangle) \tag{3.20}$$

i.e. The state is unentangled, and Alice's particle is in the state $|0\rangle$ and Bob's particle is in the (normalized version of) $\alpha_{00}|0\rangle + \alpha_{01}|1\rangle$. Had Alice got the outcome $\lambda = -1$, Alice's particle will be in the state $|1\rangle$ and Bob's particle is in the (normalized version of) $\alpha_{10}|0\rangle + \alpha_{11}|1\rangle$. We note that Alice's two states are orthogonal, but Bob's two states after the measurement may or may not be orthogonal.

More generally, consider Alice and Bob each have a qubit and Alice measures a general operator

$$A = \lambda_1|e_1\rangle\langle e_1| + \lambda_2|e_2\rangle\langle e_2| \tag{3.21}$$

where $\{|e_1\rangle, |e_2\rangle\}$ form an arbitrary orthonormal basis for $\mathbb{C}^2$, and $\lambda_1 \neq \lambda_2$. Then the operator on the whole Hilbert space is $A \otimes I$. If Alice gets the outcome λ_1, then state becomes (the normalized version of)

$$|e_1\rangle\langle e_1| \otimes I|\Psi\rangle = |e_1\rangle|\eta\rangle \tag{3.22}$$

where $|\eta\rangle$ is some state of Bob's system [i.e. the state is a product state].

We end this section with a few notational observations.

1. Firstly note we can write operators on $\mathbb{C}^2 \otimes \mathbb{C}^2$ in tensor product form or directly in terms of states on the whole Hilbert space. e.g.

$$|0\rangle\langle 0| \otimes |1\rangle\langle 1| = |01\rangle\langle 01|. \tag{3.23}$$

More generally if $a, b, c, d = 0, 1$ label the basis states, then

$$|a\rangle\langle b| \otimes |c\rangle\langle d| = |ac\rangle\langle bd|. \tag{3.24}$$

2. Matrices of operators on tensor product Hilbert spaces.

Recall from earlier in the course that given an orthonormal basis for the Hilbert space on which the operator acts, $\{|v_j\rangle\}$, the matrix of an operator A can be written as

$$A_{jk} = \langle v_j|A|v_k\rangle. \tag{3.25}$$

So consider $\mathbb{C}^2 \otimes \mathbb{C}^2$. The standard basis is $|00\rangle, |01\rangle, |10\rangle, |11\rangle$, in that order. Thus the 4x4 matrix of an operator on $\mathbb{C}^2 \otimes \mathbb{C}^2$ is

$$\begin{bmatrix} A_{00,00} & A_{00,01} & \cdots & \\ A_{01,00} & \cdots & & \\ \vdots & & & \\ A_{11,00} & \cdots & & A_{11,11} \end{bmatrix} \tag{3.26}$$

where

$$A_{00,00} = \langle 00|A|00\rangle \quad \text{etc.} \tag{3.27}$$

Thus for example the matrix of $X \otimes Z$ with respect to the standard basis is

$$\begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \\ 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{bmatrix}. \tag{3.28}$$

4 Applications

4.1 No cloning theorem

Consider a proposed universal copying operator **C**. **C** should start with a particle in an arbitrary state $|\psi\rangle$ and a particle in a "blank" state $|\Omega\rangle$ and copy the state of the first particle onto the state of the second particle, while keeping the first particle in the state $|\psi\rangle$:

$$\mathbf{C}|\psi\rangle|\Omega\rangle = |\psi\rangle|\psi\rangle. \tag{4.1}$$

Since it is assumed to be universal [i.e. can copy any state] it should also copy the state of the first particle if it is another state $|\phi\rangle$:

$$\mathbf{C}|\phi\rangle|\Omega\rangle = |\phi\rangle|\phi\rangle. \tag{4.2}$$

But since quantum operations are linear, **C** will also copy the superposition of $|\psi\rangle$ and $|\phi\rangle$ as follows, using the previous two equations:

$$\mathbf{C}\Big(|\psi\rangle + |\phi\rangle\Big)|\Omega\rangle = |\psi\rangle|\psi\rangle + |\phi\rangle|\phi\rangle. \tag{4.3}$$

But this is not what we wanted. If the cloner worked as we wanted, we should have mapped

$$\Big(|\psi\rangle + |\phi\rangle\Big)|\Omega\rangle \mapsto \Big(|\psi\rangle + |\phi\rangle\Big)\Big(|\psi\rangle + |\phi\rangle\Big). \tag{4.4}$$

Thus the cloner that correctly copies $|\psi\rangle$ and $|\phi\rangle$ does not correctly copy the superposition, and so there is no universal cloner.

4.2 Comparison between classical and quantum information

• Classical

Described by bits; a classical n-bit system can be in any one of 2^n states 000...0 to 111...1. And classical information can be written, read and copied.

• Quantum

Described by qubits, systems that can be in states of the form

$$\alpha|0\rangle + \beta|1\rangle. \tag{4.5}$$

A quantum n-qubit system can be in an arbitrary superposition of the 2^n states $|000...0\rangle$ to $|111...1\rangle$. We need infinitely many bits to describe the state of a single qubit [think of the binary expansions of the real and imaginary parts of α]. Thus we can say, informally, that we can write an infinite amount of information onto a qubit. However we cannot read this information [when making a measurement we do not learn α or β; for example if we measure Z we only get one of the two eigenvalues $+1$ or -1, and typically all memory of α or β is lost]. And finally, we cannot copy quantum information.

4.3 Super-dense coding

• Sending a bit in a qubit

Imagine that Alice can send a qubit to Bob, but she in fact wants to send bit. This is easy: for example she could encode 0 as $|0\rangle$ and 1 as $|1\rangle$, and send the qubit to Bob. Bob measures Z for example and can decide whether Alice had sent $|0\rangle$ or $|1\rangle$, with certainty [as these are eigenstates of Z with different eigenvalues] and thus learn what bit value Alice had wanted to send.

But if Alice had wanted to send two bits [i.e. one of four messages], it would not work for example for her to send one of four states in $\mathbb{C}^2$ [e.g. $|0\rangle, |1\rangle, |+\rangle, |-\rangle$] as there are not four orthogonal states in $\mathbb{C}^2$. Thus Bob cannot reliably distinguish which state Alice had sent. e.g. in the example given, if Bob measures Z and gets eigenvalue $+1$, the state Alice had sent could have been any state apart from $|1\rangle$.

Nonetheless, by sending only one qubit, Alice can send two bits to Bob, using the Super-dense coding protocol:

Handout K Super-dense coding

• Alice wants to send one of four messages to Bob (i.e. two bits), but can only send one qubit.

• She could send the qubit either in the state $|0\rangle$ or $|1\rangle$. If Bob were to measure Z, say, he learns what the state is, but this only supplies one bit (i.e. only one of two possible messages)

• However it can be done!

• Alice and Bob pre-arrange to share two particles (one held by Alice and one by Bob) in the state

$$\frac{1}{\sqrt{2}}\Big[|0\rangle_A|0\rangle_B + |1\rangle_A|1\rangle_B\Big].$$

Then Alice does one of four things to her particle before she sends it

$$\begin{aligned}
00 &\rightarrow \text{do nothing} & \frac{1}{\sqrt{2}}\Big[|0\rangle_A|0\rangle_B + |1\rangle_A|1\rangle_B\Big] &= |\Phi_I\rangle \\
01 &\rightarrow X & \frac{1}{\sqrt{2}}\Big[|1\rangle_A|0\rangle_B + |0\rangle_A|1\rangle_B\Big] &= |\Phi_X\rangle \\
10 &\rightarrow Y & \frac{1}{\sqrt{2}}\Big[|1\rangle_A|0\rangle_B - |0\rangle_A|1\rangle_B\Big] &= |\Phi_Y\rangle \\
11 &\rightarrow Z & \frac{1}{\sqrt{2}}\Big[|0\rangle_A|0\rangle_B - |1\rangle_A|1\rangle_B\Big] &= |\Phi_Z\rangle
\end{aligned}$$

Alice now sends her qubit. Bob now has two particles in one of four possible orthogonal states. If he makes a measurement of an operator which has these four states as eigenstates, he will discover *with certainty* which state he has, and so which message Alice intended to send.

• Thus Alice has managed to communicate two bits.

4.4 Teleportation

Before describing teleportation let us revisit what happens when one part of a multi-party system is measured.

Consider a general two-qubit state

$$|\psi\rangle = \alpha_{00}|00\rangle + \alpha_{01}|01\rangle + \alpha_{10}|10\rangle + \alpha_{11}|11\rangle. \tag{4.6}$$

Consider now that Alice meaures A on her particle, where A is any operator with eigenstates $|0\rangle$ and $|1\rangle$ with different eigenvalues.

$$A = \lambda_0|0\rangle\langle 0| + \lambda_1|1\rangle\langle 1| \tag{4.7}$$

This measurement is represented by

$$A \otimes I = \lambda_0|0\rangle\langle 0| \otimes I + \lambda_1|1\rangle\langle 1| \otimes I := \lambda_0 P_0 + \lambda_1 P_1 \tag{4.8}$$

on the whole Hilbert space.

We want to know what the state of the system will be after the measurement. To do so it is convenient to write the state as

$$|\psi\rangle = |0\rangle(\alpha_{00}|0\rangle + \alpha_{01}|1\rangle) + |1\rangle(\alpha_{10}|0\rangle + \alpha_{11}|1\rangle). \tag{4.9}$$

Thus we can see that if Alice gets eigenvalue λ_0 then we can compute

$$P_0|\psi\rangle = |0\rangle(\alpha_{00}|0\rangle + \alpha_{01}|1\rangle), \tag{4.10}$$

and so the state after the measurement is

$$\frac{|0\rangle(\alpha_{00}|0\rangle + \alpha_{01}|1\rangle)}{|||0\rangle(\alpha_{00}|0\rangle + \alpha_{01}|1\rangle)||} = \frac{|0\rangle(\alpha_{00}|0\rangle + \alpha_{01}|1\rangle)}{\sqrt{|\alpha_{00}|^2 + |\alpha_{01}|^2}}, \tag{4.11}$$

and the probability that this happens is

$$|\alpha_{00}|^2 + |\alpha_{01}|^2. \tag{4.12}$$

Thus it is helpful to think of $|\psi\rangle$ as

$$|\psi\rangle = |0\rangle|\eta_0\rangle + |1\rangle|\eta_1\rangle, \tag{4.13}$$

and we can read off that the state after the measurement will be [up to normalization]

$$\begin{array}{ll} & |0\rangle|\eta_0\rangle \\ \text{or} & |1\rangle|\eta_1\rangle \end{array} \tag{4.14}$$

depending whether the measurement yields λ_0 or λ_1. Notice also that if Alice had got eigenvalue λ_0 (resp. λ_1) she knows Bobs qubit is in state $|\eta_0\rangle$ (resp. $|\eta_1\rangle$) after the measurement

We can extend this to the situation where Alice has access to a four dimensional Hilbert space and Bob a qubit. Let Alice plan to measure a general non-degenerate operator

$$A_4 = \lambda_1|\phi_1\rangle\langle\phi_1| + \lambda_2|\phi_2\rangle\langle\phi_2| + \lambda_3|\phi_3\rangle\langle\phi_3| + \lambda_4|\phi_4\rangle\langle\phi_4|, \tag{4.15}$$

where $|\phi_1\rangle$, $|\phi_2\rangle$,$|\phi_3\rangle$,$|\phi_4\rangle$, the eigenstates of A_4, form an orthonormal basis for $\mathbb{C}^4$.

In this case it will be useful to write a general state on the whole system as

$$|\Psi\rangle = |\phi_1\rangle|\tau_1\rangle + |\phi_2\rangle|\tau_2\rangle + |\phi_3\rangle|\tau_3\rangle + |\phi_4\rangle|\tau_4\rangle \tag{4.16}$$

where $|\tau\rangle_1$ to $|\tau\rangle_4$ are the four single qubit states that arise when we write out $|\Psi\rangle$ in this way. These single qubit states will not all be orthogonal in general.

The advantage of writing $|\Psi\rangle$ is this way is that we can read off what will happen when Alice measures A_4. The operator on the whole Hilbert space is $A_4 \otimes I$. Thus, for example, the projector associated to the outcome λ_2 is

$$|\phi_2\rangle\langle\phi_2| \otimes I \tag{4.17}$$

So we can read off that

$$|\phi_2\rangle\langle\phi_2| \otimes I|\Psi\rangle = |\phi_2\rangle|\tau_2\rangle, \tag{4.18}$$

and so the probability of getting λ_2 is

$$|||\phi_2\rangle|\tau_2\rangle||^2. \tag{4.19}$$

And the state after the measurement is

$$\frac{|\phi_2\rangle|\tau_2\rangle}{|||\phi_2\rangle|\tau_2\rangle||}. \tag{4.20}$$

This way of writing the state $|\Psi\rangle$ will be useful when computing what happens during quantum teleportation.

Handout L Teleportation

• Alice has an unknown state $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$ which she wishes to transmit to Bob. She cannot find out what the state is with certainty, nor can she copy it. Nonetheless she can transmit it to Bob.

Initially Alice and Bob also share a maximally entangled state. Thus the complete state is

$$\Big[\alpha|0\rangle_A + \beta|1\rangle_A\Big]\frac{1}{\sqrt{2}}\Big[|0\rangle_{A_1}|0\rangle_B + |1\rangle_{A_1}|1\rangle_B\Big].$$

We may re-write this state as follows

$$\begin{aligned}
&\frac{1}{2\sqrt{2}}\Big[|0\rangle_A|0\rangle_{A_1} + |1\rangle_A|1\rangle_{A_1}\Big]\Big[\alpha|0\rangle_B + \beta|1\rangle_B\Big] \\
&+\frac{1}{2\sqrt{2}}\Big[|0\rangle_A|0\rangle_{A_1} - |1\rangle_A|1\rangle_{A_1}\Big]\Big[\alpha|0\rangle_B - \beta|1\rangle_B\Big] \\
&+\frac{1}{2\sqrt{2}}\Big[|0\rangle_A|1\rangle_{A_1} + |1\rangle_A|0\rangle_{A_1}\Big]\Big[\beta|0\rangle_B + \alpha|1\rangle_B\Big] \\
&+\frac{1}{2\sqrt{2}}\Big[|0\rangle_A|1\rangle_{A_1} - |1\rangle_A|0\rangle_{A_1}\Big]\Big[-\beta|0\rangle_B + \alpha|1\rangle_B\Big].
\end{aligned}$$

Let Alice measure an operator on her four-dimensional Hilbert space with four un-equal eigenvalues $\lambda_1, \lambda_2, \lambda_3, \lambda_4$, with associated eigenvectors

$$\begin{aligned}
&\frac{1}{\sqrt{2}}\Big[|0\rangle_A|0\rangle_{A_1} + |1\rangle_A|1\rangle_{A_1}\Big] \quad (= |\Phi_I\rangle)\,; \\
&\frac{1}{\sqrt{2}}\Big[|0\rangle_A|0\rangle_{A_1} - |1\rangle_A|1\rangle_{A_1}\Big] \quad (= |\Phi_Z\rangle)\,; \\
&\frac{1}{\sqrt{2}}\Big[|0\rangle_A|1\rangle_{A_1} + |1\rangle_A|0\rangle_{A_1}\Big] \quad (= |\Phi_X\rangle)\,; \\
&\frac{1}{\sqrt{2}}\Big[|0\rangle_A|1\rangle_{A_1} - |1\rangle_A|0\rangle_{A_1}\Big] \quad (= -|\Phi_Y\rangle)\,.
\end{aligned}$$

Notice that the same Bell states appear in Teleportation as we used in Super-dense coding.

Depending on which outcome she finds, she knows what Bob's state must be:

$$\begin{aligned}
\lambda_1 &\leftrightarrow \Big[\alpha|0\rangle_B + \beta|1\rangle_B\Big] \\
\lambda_2 &\leftrightarrow \Big[\alpha|0\rangle_B - \beta|1\rangle_B\Big] \\
\lambda_3 &\leftrightarrow \Big[\beta|0\rangle_B + \alpha|1\rangle_B\Big] \\
\lambda_4 &\leftrightarrow \Big[-\beta|0\rangle_B + \alpha|1\rangle_B\Big].
\end{aligned}$$

Depending on which outcome she finds, she transmits one of four messages to Bob:

$$\begin{aligned}
\lambda_1 &\quad \text{do nothing} \\
\lambda_2 &\quad \text{act with} \quad Z \\
\lambda_3 &\quad \text{act with} \quad X \\
\lambda_4 &\quad \text{act with} \quad Y.
\end{aligned}$$

In all cases the Bob's state becomes $|\psi\rangle$.

Thus with initial shared entanglement and the transmission of two classical bits, the state $|\psi\rangle$ will be transmitted to Bob (without either of them knowing what the state was at any time).

A final point to make about teleportation is that the state does not appear with Bob instantaneously, but rather he has to wait until the two bits have arrived, so there is no transmission faster than the speed of light.

4.5 Quantum Computation - Introduction

We start with a number of preliminaries

• **Hadamard operator**

H is the operator on $\mathbb{C}^2$ defined by

$$H|0\rangle = |+\rangle \qquad H|1\rangle = |-\rangle. \tag{4.21}$$

Thus $H^2 = I$ and note also, on $\mathbb{C}^2 \otimes \mathbb{C}^2$

$$H \otimes H \, |0\rangle|0\rangle = \frac{1}{2}\Big(|00\rangle + |01\rangle + |10\rangle + |11\rangle\Big) \tag{4.22}$$

and when acting on the Hilbert space of n qubits,

$$H \otimes H \otimes H \ldots \otimes H|0\rangle|0\rangle|0\rangle \ldots |0\rangle = \frac{1}{\sqrt{2^n}} \sum_{\text{all } x} |x\rangle \tag{4.23}$$

• **Measuring in the computational basis** The operator C on $\mathbb{C}^2$ defined by

$$C = |1\rangle\langle 1| \tag{4.24}$$

has eigenvalues 0 and 1, corresponding to $|0\rangle$ and $|1\rangle$. If the state is one of the basis states $|0\rangle$ or $|1\rangle$, then if C is measured, the eigenvalue obtained is 0 or 1 respectively. Thus C "reads out" which of the two basis states we have.

Note that it may be helpful to think of C in its spectral decomposition as

$$C = 0|0\rangle\langle 0| + 1|1\rangle\langle 1|. \tag{4.25}$$

If we have a two-qubit state

$$|\Psi\rangle = \sum_{j,k=0}^{1} \alpha_{jk}|j\rangle|k\rangle \tag{4.26}$$

then if we measure both qubits with C, then the probability that we get 0 for the first qubit and 0 for the second qubit is $|\alpha_{00}|^2$, since, using the standard quantum mechanical rules

$$\text{Prob}\Big(0 \text{ on first qubit}\Big) = |||0\rangle\langle 0| \otimes I|\Psi\rangle||^2 := q_1, \tag{4.27}$$

and the state after the measurement is we get the eigenvalue 0 when measuring C on the first qubit is

$$\frac{|0\rangle\langle 0| \otimes I|\Psi\rangle}{|||0\rangle\langle 0| \otimes I|\Psi\rangle||}. \tag{4.28}$$

Thus

$$\begin{aligned} & \text{Prob}\Big(0 \text{ on second qubit given 0 on first qubit}\Big) \\ = \; & ||I \otimes |0\rangle\langle 0|\Big(\frac{|0\rangle\langle 0| \otimes I|\Psi\rangle}{|||0\rangle\langle 0| \otimes I|\Psi\rangle||}\Big)||^2 \\ = \; & \frac{|||0\rangle\langle 0| \otimes |0\rangle\langle 0||\Psi\rangle||^2}{|||0\rangle\langle 0| \otimes I|\Psi\rangle||^2} := q_2. \end{aligned} \tag{4.29}$$

Thus the probability that we get 0 on the first qubit and 0 on the second qubit is

$$q_1 q_2 = |||0\rangle\langle 0| \otimes |0\rangle\langle 0||\Psi\rangle||^2 = |\alpha_{00}|^2. \tag{4.30}$$

• **Classical functions** $\{0,1\}^n \to \{0,1\}$ An example in $n = 2$ is f_1 defined by:

$$f_1(00) = 1,\ f_1(01) = 0,\ f_1(10) = 1,\ f_1(11) = 0. \tag{4.31}$$

Can write this as $f_1(x_1, x_2) = 1 + x_2 \bmod 2$. We also denote this $f_1(x_1, x_2) = 1 \oplus x_2$.

• **Computing functions in parallel** Given an $f : \{0,1\}^n \to \{0,1\}$ we can construct a unitary operator acting on $n+1$ qubits:

$$U_f|x\rangle|y\rangle := |x\rangle|y \oplus f(x)\rangle. \tag{4.32}$$

$|x\rangle$ denotes one of the 2^n basis states for $\mathbb{C}^{2^n}$ [i.e. $|00\ldots0\rangle$ to $|11\ldots1\rangle$], and $|y\rangle$ is one of the two basis states for $\mathbb{C}^2$ [i.e. $|0\rangle$ or $|1\rangle$].

e.g. for the classical function f_1 above,

$$U_{f_1}|00\rangle|0\rangle = |00\rangle|1\rangle,\ U_{f_1}|00\rangle|1\rangle = |00\rangle|0\rangle,\ U_{f_1}|11\rangle|1\rangle = |11\rangle|1\rangle. \text{ etc..} \tag{4.33}$$

We can see that U_f takes an orthonormal basis to an orthonormal basis by computing the inner product of $U_f|x_2\rangle|y_2\rangle$ with $U_f|x_1\rangle|y_1\rangle$:

$$\begin{aligned}
\Big(\langle x_2|\langle y_2 \oplus f(x_2)|\Big)\Big(|x_1\rangle|y_1 \oplus f(x_1)\rangle\Big) &= \delta_{x_1x_2}\Big(\langle y_2 \oplus f(x_2)|\Big)\Big(|y_1 \oplus f(x_1)\rangle\Big) \\
&= \delta_{x_1x_2}\Big(\langle y_2 \oplus f(x_1)|\Big)\Big(|y_1 \oplus f(x_1)\rangle\Big) \\
&= \delta_{x_1x_2}\langle y_2|y_1\rangle \\
&= \delta_{x_1x_2}\delta_{y_1y_2} \\
&= \Big(\langle x_2|\langle y_2|\Big)\Big(x_1\rangle|y_1\rangle\Big),
\end{aligned} \tag{4.34}$$

as required.

We note that any operator of the form

$$\sum_{j=0}^{d-1} |e_j\rangle\langle f_j|, \tag{4.35}$$

where $\{|e_j\rangle\}$ and $\{|f_j\rangle\}$ are two orthonormal bases of $\mathbb{C}^d$, is unitary [as can be checked by computing $U^\dagger U$]. The proof above that U_f takes an orthonormal basis to an orthonormal basis shows that U_f is indeed of this form, and so is unitary.

A key property of U_f is that it allows us to compute functions in parallel [i.e. we can compute the function on many or all values of the input x simultaneously]:

$$U_f\Big(\frac{1}{\sqrt{2^n}}\sum_x |x\rangle\Big)|0\rangle = \Big(\frac{1}{\sqrt{2^n}}\sum_x |x\rangle|f(x)\rangle\Big). \tag{4.36}$$

Finally we note

$$\begin{aligned}
U_f|x\rangle|-\rangle &= U_f|x\rangle\Big(\frac{|0\rangle - |1\rangle}{\sqrt{2}}\Big) \\
&= \frac{1}{\sqrt{2}}|x\rangle\Big(|f(x)\rangle - |f(x) \oplus 1\rangle\Big) \\
&= = \begin{cases} |x\rangle|-\rangle & \text{if } f(x) = 0 \\ -|x\rangle|-\rangle & \text{if } f(x) = 1 \end{cases} \\
&= (-1)^{f(x)}|x\rangle|-\rangle.
\end{aligned} \tag{4.37}$$

• **Trits** A *trit* is an object that can be in one of three states $\{0,1,2\}$. And we can generalise the ideas concerning bits to trits.

Thus we can think of functions from trits to trits; $f : \{0,1,2\} \mapsto \{0,1,2\}$ such as:

$$\begin{aligned} f_1(0) &= 1, \quad f_1(1) = 1, \quad f_1(2) = 1 \\ f_2(0) &= 0, \quad f_2(1) = 1, \quad f_2(2) = 1 \\ f_3(0) &= 2, \quad f_3(1) = 0, \quad f_3(2) = 0. \end{aligned} \tag{4.38}$$

We can write these functions as follows:

$$f_1(x) = 1, \quad f_2(x) = x^2 \bmod 3, \quad f_3(x) = x^2 + 2 \bmod 3, \tag{4.39}$$

where "mod 3" denotes the value modulo 3.

Consider also the function $f_4(x) = 3x^2 \bmod 3$. Because of the "mod 3", in fact this function is $f_4(x) = 0$.

Just as with bits to bits, we can associate a unitary operator with classical functions from trits to trits. It is an operator on $\mathbb{C}^3 \otimes \mathbb{C}^3$ defined as follows

$$U_f|x\rangle|y\rangle = |x\rangle|y + f(x) \bmod 3\rangle, \tag{4.40}$$

where $x \in \{0,1,2\}$ and $y \in \{0,1,2\}$.

4.6 Deutsch's problem

The problem is: given a black box for an unknown function $f : \{0,1\} \mapsto \{0,1\}$, compute $f(0) \oplus f(1)$.

The four functions are

	$f_1(x)$	$f_2(x)$	$f_3(x)$	$f_4(x)$
$x = 0$	0	0	1	1
$x = 1$	0	1	0	1
$f(0) \oplus f(1)$	0	1	1	0

Two queries to a classical black box are needed. For example, if 0 is inserted into the black box and the box returns 0, then the function could have been either f_1 or f_2 since $f_1(0) = f_2(0) = 0$. But $f_1(0) \oplus f_1(1) = 0$ whereas $f_2(0) \oplus f_2(1) = 1$ so we need to insert 1 into the box as well in order to learn $f(0) \oplus f(1)$. Similarly whatever single input is put into the black box, and whatever outcome is returned, there are two possible functions consistent with that single input and the functions have different values for $f(0) \oplus f(1)$. Thus inserting one input into the black box is not enough to learn $f(0) \oplus f(1)$, and both 0 and 1 have to be inserted into the black box. i.e. we need two uses of the box, in other words two queries.

Handout M Deutsch's Algorithm

The aim is to find out whether $f(0) \oplus f(1) = 0$ or $f(0) \oplus f(1) = 1$

The algorithm

1. Put 2 qubits into the state $|0\rangle|0\rangle$

2. Apply $I \otimes X$

3. Apply $H \otimes H$

4. Apply U_f

5. Apply $H \otimes I$

6. Measure first qubit in the computational basis.

If the value is 0 then $f(0) \oplus f(1) = 0$. If the value is 1 then $f(0) \oplus f(1) = 1$.

Analysis of the algorithm

1. Put 2 qubits into the state

$$|0\rangle|0\rangle$$

2. Apply $I \otimes X$

State becomes

$$|0\rangle|1\rangle$$

3. Apply $H \otimes H$

State becomes

$$\left(\frac{|0\rangle + |1\rangle}{\sqrt{2}}\right)\left(\frac{|0\rangle - |1\rangle}{\sqrt{2}}\right)$$

4. Apply U_f

Recall that

$$U_f|x\rangle\left(\frac{|0\rangle - |1\rangle}{\sqrt{2}}\right) = (-1)^{f(x)}|x\rangle\left(\frac{|0\rangle - |1\rangle}{\sqrt{2}}\right)$$

Thus the state becomes

$$\left(\frac{(-1)^{f(0)}|0\rangle + (-1)^{f(1)}|1\rangle}{\sqrt{2}}\right)\left(\frac{|0\rangle - |1\rangle}{\sqrt{2}}\right)$$

$$= \begin{cases} \pm|+\rangle|-\rangle & \text{if} \quad f(0)\oplus f(1) = 0 \\ \pm|-\rangle|-\rangle & \text{if} \quad f(0)\oplus f(1) = 1 \end{cases}$$

where as usual

$$|\pm\rangle = \frac{|0\rangle \pm |1\rangle}{\sqrt{2}}$$

Thus the two cases $f(0)\oplus f(1) = 0$ and $f(0)\oplus f(1) = 1$ can be distinguised by a measurement. But this measurement should be in the computational basis.

5. Apply $H \otimes I$

State becomes

$$= \begin{cases} \pm|0\rangle|-\rangle & \text{if} \quad f(0)\oplus f(1) = 0 \\ \pm|1\rangle|-\rangle & \text{if} \quad f(0)\oplus f(1) = 1 \end{cases}$$

6. Measure first qubit in the computational basis.

The value of the outcome is $f(0)\oplus f(1)$.

4.7 The Deutsch-Jozsa problem

Deutsch's problem can be generalized to give an exponential separation between the number of queries on a classical computer versus the number on a quantum computer.

We now consider functions $f : \{0,1\}^n \mapsto \{0,1\}$. We are given the "promise" that the function is constant [the same value on all inputs] or "balanced" [0 on exactly half the inputs 1 on the other half]. The DJ problem is that we have a black box for an unknown function that is promised to be either constant or balanced. We only wish to determine which of the two types it is [and we want to know this with certainty].

e.g. $n = 2$ Here are some examples of functions $f : \{0,1\}^2 \mapsto \{0,1\}$

	$f_0(x)$	$f_1(x)$	$f_2(x)$	$f_3(x)$
$x = 00$	1	0	1	1
$x = 01$	1	1	1	0
$x = 10$	1	1	0	0
$x = 11$	1	0	0	0
	constant	balanced	balanced	neither

For $n = 2$ there are two constant functions [$f(x) = 0\ \forall x$ and $f(x) = 1\ \forall x$] and six balanced functions [the two 0 values can be the values on any pair of inputs].

If we have a classical black box for the function, $2^{n-1} + 1$ queries are needed in the worst case. For consider inputting 2^{n-1} different values x; if it turns out that they all have $f(x) = 0$, say, then we still don't know whether the function is constant or balanced.

• **The quantum algorithm for the Deutsch-Jozsa problem**

Only one use of U_f is needed, as described in Handout N. Thus there is an exponential separation between the number of queries needed classically ($2^{n-1} + 1$) and quantum mechanically (one).

Handout N The Deutsch-Jozsa Algorithm

We are given a quantum black box that is associated to a classical function $f : \{0,1\}^n \to \{0,1\}$ that is promised to be be balanced or constant. We want to find out which.

The algorithm

1. Put $n+1$ qubits into the state $|0\rangle^{\otimes n}|0\rangle$

2. Apply $I^{\otimes n} \otimes X$

3. Apply $H^{\otimes n} \otimes H$

4. Apply U_f

5. Apply $H^{\otimes n} \otimes I$

6. Measure first n qubits in the computational basis.

If all values are 0 then the function f was constant. Otherwise the function was balanced.

Analysis of the algorithm

1. Put $n+1$ qubits into the state

$$|0\rangle^{\otimes n}|0\rangle$$

2. Apply $I^{\otimes n} \otimes X$

State becomes

$$|0\rangle^{\otimes n}|1\rangle$$

3. Apply $H^{\otimes n} \otimes H$

State becomes

$$\left(\frac{|0\rangle+|1\rangle}{\sqrt{2}}\right)\left(\frac{|0\rangle+|1\rangle}{\sqrt{2}}\right)\cdots\left(\frac{|0\rangle+|1\rangle}{\sqrt{2}}\right)\left(\frac{|0\rangle-|1\rangle}{\sqrt{2}}\right)$$
$$=\frac{1}{2^{n/2}}\sum_x|x\rangle\left(\frac{|0\rangle-|1\rangle}{\sqrt{2}}\right)$$

4. Apply U_f

State becomes

$$\frac{1}{2^{n/2}}\sum_x(-1)^{f(x)}|x\rangle\left(\frac{|0\rangle-|1\rangle}{\sqrt{2}}\right)$$

Note if we call

$$|f\rangle=\frac{1}{2^{n/2}}\sum_x(-1)^{f(x)}|x\rangle$$

then the states $|f\rangle$ for constant functions are all orthogonal to the those for balanced functions. Thus there will be a way of distinguishing between the two sets by a measurement. But this measurement should be in the computational basis.

5. Apply $H^{\otimes n}\otimes I$

Note that if f is constant, then

$$H^{\otimes n}|f\rangle=\pm|0\rangle^{\otimes n}$$

since H is self-inverse.

And so if f is balanced, then

$$H^{\otimes n}|f\rangle=\sum_x a_x|x\rangle$$

must have

$$a_{00\ldots0}=0$$

6. Measure first n qubits in the computational basis.

If all values are 0 then the function f was constant. Otherwise the function was balanced: we cannot get 0 for every measurement if f was balanced - we must get at least one 1 outcome on one of the qubits. This is because the probability of getting the string x is $|a_x|^2$.

4.8 Quantum Cryptography

The aim is to send a message from the sender to the receiver in such a way that an un-authorized recipient should not be able to read the message.

Some coding schemes use mathematical techniques to encode.

An early example is the Caesar cipher. Here each letter of the alphabet in the text to be sent is replaced by one three places further in the alphabet [i.e. A is replaced by D, B by E, X replaced with A etc.]

Thus Alice might send the following secret message to Bob:

LORYHBRXERE.

The problem with such encoding is that if an eavesdropper suspects that such a replacement is being made, for example with a letter being replaced by a letter k places later in the alphabet, not many trials [i.e. 25] of different possible k's allow the eavesdropper to work out the message.

It might be thought that a better strategy is to replace every letter by applying a random permutation to the set of letters. Now the number of possible strategies is $26! \simeq 2^{88} \simeq 10^{26}$, so it is impossible to try all different possibilities in a reasonable time. However it turns out this is not a good strategy for anything other than very short texts, since it is a perhaps surprising fact about almost all English text that the letter E will be most common followed by T and so on. So we can use "letter frequency" to make a good attempt at decoding text encoded using a random permutation.

These two schemes [and many others that are used at present] are of the type whose diffculty of being cracked [or otherwise] depend on the difficulty of solving certain types of mathematical problem.

A very different, and often safer, strategy is to use shared random strings to encode data.

The classical and quantum versions of this key distribution are described in Handout O. A key point about quantum key distribution is that its security does not depend on the mathematical difficulty of a problem, but rather on the laws of Physics.

Handout O Key Distribution

CLASSICAL ONE TIME PAD

Alice wants to send a message secretly to Bob - think of it as a binary string:

$$\mathbf{a} = 0\ \ 1\ \ 1\ \ 0\ \ 1\ \ 1\ \ 0\ \ 1\ \ 0\ \ 1 \quad \textit{Alice's message}$$

Consider that they previously arrange to share, secretly, a long random string of bits - the key $\mathbf{k}$. Often this is done in practice by a person physically taking the key - a "courier"

$$\mathbf{k} = 1\ \ 0\ \ 1\ \ 0\ \ 1\ \ 1\ \ 0\ \ 0\ \ 0\ \ 0 \quad \textit{shared secret key}$$

The message Alice sends to Bob $\mathbf{m}$ is the bit-wise sum of these

$$\begin{array}{lllllllllll} \mathbf{a} & = 0 & 1 & 1 & 0 & 1 & 1 & 0 & 1 & 0 & 1 & \textit{Alice's message} \\ \mathbf{k} & = 1 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & \textit{shared secret key} \\ \mathbf{m} = \mathbf{a} \oplus \mathbf{k} & = 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & \textit{message} \end{array}$$

The key point is that $\mathbf{m}$ is a random string, so an eavesdropper, Eve, since she is assumed not to know $\mathbf{k}$, can learn nothing about $\mathbf{a}$ from it. However Bob can easily recover $\mathbf{a}$ by adding $\mathbf{k}$ to $\mathbf{m}$:

$$\begin{array}{lllllllllll} \mathbf{a} & = 0 & 1 & 1 & 0 & 1 & 1 & 0 & 1 & 0 & 1 & \textit{Alice's message} \\ \mathbf{k} & = 1 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & \textit{shared secret key} \\ \mathbf{m} = \mathbf{a} \oplus \mathbf{k} & = 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & \textit{message} \\ \mathbf{m} \oplus \mathbf{k} & = 0 & 1 & 1 & 0 & 1 & 1 & 0 & 1 & 0 & 1 & \textit{Bob's decoded message} \end{array}$$

Notes

- This all works fine if Eve does not know $\mathbf{k}$. But since this is a classical bit string, Eve might be able to get access to it, and copy it.
- If Eve has intercepted $\mathbf{k}$ and Alice and Bob learn that she has, then they can no longer use the one time pad. Similarly in the case of ciphers such as the Caesar cipher, if Eve knows how the cipher works, then she can read all messages sent between Alice and Bob
- It is a particular problem if Eve has managed to do this without Alice and Bob knowing since they then continue to send messages assuming that they are secure. Classically this is possible.
- However if Alice and Bob have a shared random key, unknown to Eve, then they have security.

QUANTUM KEY DISTRIBUTION

This is a method for providing Alice and Bob a shared, random secret key [NB this is the key to used later for sending a message, not the message to be sent]. But in particular, if Eve has intercepted the communication, it allows us to learn that she has done so.

INTUITION

(a) Cannot make measurement of system without disturbing it. So if eavesdropper, Eve, intercepts message, she will disturb it.

e.g. If state is

$$\alpha|0\rangle + \beta|1\rangle,$$

and if a measurement of Z is made, then one does not learn what α or β were, and after the measurement, the state becomes $|0\rangle$ or $|1\rangle$.

(b) Cannot reliably distinguish two non-orthogonal states. e.g.

$$|+\rangle = \frac{|0\rangle + |1\rangle}{\sqrt{2}} \quad \text{and} \quad |0\rangle$$

(c) Cannot copy a quantum state [no cloning theorem].

Thus unlike a classical one-time pad, a treacherous courier cannot copy a quantum state and hand it on.

Bad protocol for producing shared random key

Alice produces random string of 0's and 1's.

She encodes 0 as $|0\rangle$ and 1 as $|1\rangle$ and sends the string of qubits to Bob.

Bob measures Z on each qubit and if he gets $\lambda = +1$ he interprets this as 0 and if he gets $\lambda = -1$ he interprets it as 1. He now shares the random key with Alice.

The problem

Eve could intercept each qubit and measure Z on it. She will learn what state the qubit is in without changing the state, so she can send the qubit on to Bob without him knowing that it was intercepted. Thus Eve can learn the key without Alice and Bob realizing.

BENNETT 92 (B92)

INTRODUCTION

The aim is for Alice and Bob to share a secret random string.

Consider the situation in which Alice sends Bob a qubit either in state $|0\rangle$ or $|+\rangle$.
If Bob measures Z and gets $\lambda = +1$ he is not sure which state Alice sent, it could have been $|0\rangle$ or $|+\rangle$, but if he measures Z and gets $\lambda = -1$, he is sure Alice sent $|+\rangle$
- he could not get $\lambda = -1$ if the state on which he makes the measurement was $|0\rangle$.
Similarly, if Bob measures X and gets $\lambda = +1$ he is not sure which state Alice sent, it could have been $|0\rangle$ or $|+\rangle$, but if he measures X and gets $\lambda = -1$, he is sure Alice sent $|0\rangle$.

THE PROTOCOL

STEP 1

Alice chooses a random string:

1	1	0	0	1	0	1	1	0	0	*Alice's random string*

She encodes "0" as $|0\rangle$ and "1" as $|+\rangle$:

1	1	0	0	1	0	1	1	0	0	*Alice's random string*
$\|+\rangle$	$\|+\rangle$	$\|0\rangle$	$\|0\rangle$	$\|+\rangle$	$\|0\rangle$	$\|+\rangle$	$\|+\rangle$	$\|0\rangle$	$\|0\rangle$	*encoding*

She now sends the qubits to Bob.

STEP 2

Bob chooses his own random bit string [independent of Alice's choice]. He will use this to decide which operator to measure on each of the qubits. If he got "0" he chooses to measure X and if he got "1" Bob chooses to measure Z:

1	1	0	0	1	0	1	1	0	0	*Alice's random string*
$\|+\rangle$	$\|+\rangle$	$\|0\rangle$	$\|0\rangle$	$\|+\rangle$	$\|0\rangle$	$\|+\rangle$	$\|+\rangle$	$\|0\rangle$	$\|0\rangle$	*encoding*
1	0	1	0	0	1	1	1	0	0	*Bob's random string*
Z	X	Z	X	X	Z	Z	Z	X	X	*Bob's measurement*

He looks at those cases where he got $\lambda = -1$. In those cases, he takes his bit to be "0" if he has measured X and "1" if he has measured Z.

1	1	0	0	1	0	1	1	0	0	*Alice's random string*
$\|+\rangle$	$\|+\rangle$	$\|0\rangle$	$\|0\rangle$	$\|+\rangle$	$\|0\rangle$	$\|+\rangle$	$\|+\rangle$	$\|0\rangle$	$\|0\rangle$	*encoding*
1	0	1	0	0	1	1	1	0	0	*Bob's random string*
Z	X	Z	X	X	Z	Z	Z	X	X	*Bob's measurement*
-1	$+1$	$+1$	$+1$	$+1$	$+1$	-1	-1	$+1$	-1	*measurement outcome*
✓	?	?	?	?	?	✓	✓	?	✓	*outcome is minus?*
1	?	?	?	?	?	1	1	?	0	*Bob's bit*

STEP 3

Bob announces publicly the cases in which he was sure which state Alice sent [i.e. those cases when he got the eigenvalue -1 when measuring X or Z]. He does not announce whether he measured X or Z! Alice and Bob discard the other cases, and retain the shared random string [In this case 1110].

STEP 4

Alice and Bob check that there has been no interference in creation of the key by selecting a random subset of the final string. They announce publicly which bits they are and what the values were. Alice and Bob should have the same values. If the values are different of some of these check bits, they can deduce that there may be a problem in the security of the key. NB they discard the check bits and do not use them for the key.

SECURITY

Eve would like to intercept qubit, copy it then send it on - then she would have same qubit as Bob. She could then wait until Alice and Bob's announcements and decide which bits to keep. [*"intercept and resend"*]. However, she cannot reliably read the qubit or copy it.

We will not do a complete security analysis.

We consider that Eve knows that Bob uses X and Z measurements, and that Alice sends qubits in state $|0\rangle$ or $|+\rangle$.

Imagine Alice sends $|0\rangle$ and Eve measures X. She can get $\lambda = -1$ or $\lambda = +1$ with equal probability.

In the case $\lambda = -1$ Eve knows that Alice has sent $|0\rangle$ and Eve resends $|0\rangle$. In this case Eve has learnt what state was sent without disturbing anything.

In the case $\lambda = +1$ Eve does not know whether Alice sent $|0\rangle$ or $|+\rangle$. Eve needs to guess what to send to Bob. If Eve decides to resend $|0\rangle$, she has guessed right and not disturbed anything.

But if she guesses wrong and sends $|+\rangle$, Bob has a different state from what Alice sent.

This mistake will sometimes be clear to Alice and Bob. For example if Bob happened to measure Z on this qubit, he could get $\lambda = -1$. Then according to the protocol, Alice and Bob would keep this bit and Bob would interpret it as 1 whereas Alice's bit was 0.

Thus if Alice and Bob were to choose a subset of their final strings and compare them, they would find that sometimes their strings are different.

Key point is that the security of the protocol follows from the laws of physics, rather than the difficulty in solving some mathematical problem ("code-breaking").

5 Density Operators

5.1 The density operator

Consider a source that emits quantum states at random. For example, the source might toss a biased coin, and depending upon the outcome observed, emit a quantum system in either the state $|0\rangle$ or $|+\rangle$, with probabilities $\frac{1}{4}$ and $\frac{3}{4}$ respectively. Such sources can definitely occur in nature. The question is how to characterise the properties of the emitted system? We will see that we can do this using the so-called ***density operator formalism.***

Consider measuring the operator $Z = |0\rangle\langle 0| - |1\rangle\langle 1|$ on the emitted system. The probability to obtain the outcome $+1$, using the standard rules of probability, is just the average probability to do so, given the state,

$$\begin{aligned}\text{Prob}(+1) &= \text{Prob}(\text{state is } |0\rangle)\,\text{Prob}(+1|\text{ state is } |0\rangle) + \text{Prob}(\text{state is } |+\rangle)\,\text{Prob}(+1|\text{ state is } |+\rangle)\\ &= \frac{1}{4}\left(\||0\rangle\langle 0|)\,|0\rangle\right\|^2 + \frac{3}{4}\left(\||0\rangle\langle 0|)\,|+\rangle\right\|^2\\ &= \frac{1}{4}\langle 0|0\rangle\langle 0|0\rangle + \frac{3}{4}\langle 0|+\rangle\langle +|0\rangle\\ &= \frac{5}{8}.\end{aligned} \tag{5.1}$$

Although this is perfectly correct, what we are looking for is a more elegant or powerful way of arriving at this result. We can begin to see how to do so, by noticing that we can rewrite the third line above as

$$\text{Prob}(+1) = \langle 0|\left(\frac{1}{4}|0\rangle\langle 0| + \frac{3}{4}|+\rangle\langle +|\right)|0\rangle. \tag{5.2}$$

This equation is interesting as we have formed an operator inside the bracket which ***depends only upon the source***. This operator is known as the ***density operator***, and is very frequently denoted by ρ,

$$\rho = \frac{1}{4}|0\rangle\langle 0| + \frac{3}{4}|+\rangle\langle +|. \tag{5.3}$$

With this in place, the probability to obtain the outcome $+1$ when measuring Z is given simply by

$$\text{Prob}(+1) = \langle 0|\rho|0\rangle. \tag{5.4}$$

There was nothing special in the above about the specific example we considered – either in terms of the source, or the measurement. More generally, suppose that the source emits systems in one of the N states $|\psi_n\rangle$, where $n \in \{1, 2, \ldots, N\}$ labels the state, with corresponding probability p_n, where $\sum_n p_n = 1$. The collection of states and associated probabilities $\{|\psi_n\rangle, p_n\}_n$ are referred to as an ***ensemble*** of states. In exact analogy to above, the density operator of this source is

$$\rho = \sum_n p_n |\psi_n\rangle\langle\psi_n|. \tag{5.5}$$

This is just the ***average*** of the projectors $|\psi_n\rangle\langle\psi_n|$ onto the states $|\psi_n\rangle$ emitted by the source.

If we now consider measuring an arbitrary operator A on the emitted system, where

$$A = \sum_k \lambda_k |e_k\rangle\langle e_k|, \tag{5.6}$$

then, in exact analogy to above, the probability to obtain the outcome λ_k is given by

$$\begin{aligned}\text{Prob}(\lambda_k) &= \sum_n \text{Prob}(\text{state is } |\psi_n\rangle)\,\text{Prob}(\lambda_k|\text{ state is } |\psi_n\rangle)\\ &= \langle e_k|\rho|e_k\rangle\end{aligned} \tag{5.7}$$

We can in fact go further, but in order to do so, we will need to introduce an important new mathematical operation, called the ***trace***. We introduce this now, and will come back to it at the end of this section, to study its properties.

The trace

The trace is the *linear* operation defined by the relation

$$\mathrm{tr}(|u\rangle\langle v|) = \langle v|u\rangle \quad \text{for all } |u\rangle, |v\rangle. \tag{5.8}$$

The fact that the trace is linear means that

$$\begin{aligned}\mathrm{tr}(\alpha|u\rangle\langle v| + \beta|w\rangle\langle x|) &= \alpha\,\mathrm{tr}(|u\rangle\langle v|) + \beta\,\mathrm{tr}(|w\rangle\langle x|),\\ &= \alpha\langle v|u\rangle + \beta\langle x|w\rangle,\end{aligned} \tag{5.9}$$

for all $|u\rangle, |v\rangle, |w\rangle, |x\rangle$ and all $\alpha, \beta \in \mathbb{C}$. One can think of the trace as 'moving' the bra $|v\rangle$ from behind the ket $|u\rangle$ to in front of it, and then contracting them together to perform the inner product.

We can use the trace to re-express (5.7), by noting that $\rho|e_k\rangle$ is mathematically a ket vector. Using (5.8) we can thus can write $\langle e_k|\,(\rho|e_k\rangle)$ instead as $\mathrm{tr}\,(|e_k\rangle\,(\langle e_k|\rho))$, and hence

$$\mathrm{Prob}(\lambda_k) = \mathrm{tr}(\rho|e_k\rangle\langle e_k|). \tag{5.10}$$

Finally, we can also recall that $P_k = |e_k\rangle\langle e_k|$ is the projection operator associated to $|e_k\rangle$, and hence

$$\mathrm{Prob}(\lambda_k) = \mathrm{tr}(\rho P_k). \tag{5.11}$$

This is the final, simplest way to express the probability of the outcome of a measurement using the density operator formalism. It says that ***the probability of the outcome of a measurement is given by the trace of the product of the density operator and the projection operator corresponding to the given outcome.***

We will now highlight and summarise some key aspects of a density operator.

5.2 Properties of the density operator

- We say that a density operator ρ is a ***mixture*** of the pure states $|\psi_n\rangle$, where each state is 'mixed in' with probability p_n.
- As seen in the original example, the states $|\psi_n\rangle$ *need not be orthogonal states*. In particular, this means that we can mix states which themselves are in ***superpositions***. In the example above, $|+\rangle$ is an equal superposition of $|0\rangle$ and $|1\rangle$, and there is no problem whatsoever in mixing this with $|0\rangle$ in a density operator. This highlights a very important point – ***the mixing that occurs in a density operator is completely distinct from quantum superposition that you are familiar with.***
- The ***number of states*** being mixed into a density operator is ***arbitrary***. It could be as few as one state (which we'll come back to shortly), or it could be an arbitrarily large number of states.
- Remarkably, ***different sources can lead to the same density operator***! In particular, it is possible that the two different ensembles $\{|\psi_n\rangle, p_n\}$ with $n \in \{1, \ldots, N\}$ and $\{|\phi_m\rangle, q_m\}$ with $m \in \{1, \ldots, M\}$, corresponding to two different sources, lead to the same density operator,

$$\sum_{n=1}^{N} p_n|\psi_n\rangle\langle\psi_n| = \sum_{m=1}^{M} q_m|\phi_m\rangle\langle\phi_m|. \tag{5.12}$$

To make this more explicit, consider the following example:

Example Consider a source which emits the eigenstates of Z, $|0\rangle$ and $|1\rangle$, with equal probability. The density operator describing the source is

$$\rho = \frac{1}{2}|0\rangle\langle 0| + \frac{1}{2}|1\rangle\langle 1| \tag{5.13}$$

Consider now a second source, which emits the eigenstates of X, $|+\rangle$ and $|-\rangle$, with equal probability. The density operator describing this source is

$$\begin{aligned}\rho &= \frac{1}{2}|+\rangle\langle +| + \frac{1}{2}|-\rangle\langle -|, \\ &= \frac{1}{2}\left(\frac{|0\rangle + |1\rangle}{\sqrt{2}}\right)\left(\frac{\langle 0| + \langle 1|}{\sqrt{2}}\right) + \frac{1}{2}\left(\frac{|0\rangle - |1\rangle}{\sqrt{2}}\right)\left(\frac{\langle 0| - \langle 1|}{\sqrt{2}}\right) \\ &= \frac{1}{4}|0\rangle\langle 0| + \frac{1}{4}|0\rangle\langle 1| + \frac{1}{4}|1\rangle\langle 0| + \frac{1}{4}|1\rangle\langle 1| + \frac{1}{4}|0\rangle\langle 0| - \frac{1}{4}|0\rangle\langle 1| - \frac{1}{4}|1\rangle\langle 0| + \frac{1}{4}|1\rangle\langle 1|, \\ &= \frac{1}{2}|0\rangle\langle 0| + \frac{1}{2}|1\rangle\langle 1|.\end{aligned} \tag{5.14}$$

This is identical to the density operator in (5.13), which arose from the source which emits $|0\rangle$ and $|1\rangle$ with equal probability.

- Crucially, we saw above that if we make any measurement, the probabilities of the outcomes depend ***only on the density operator of the source***. This has a very important implication: two sources which lead to the same density operator are ***completely indistinguishable*** – there is nothing that can be measured to figure out which source created the density operator. We will see later, that this fact is indispensable from the perspective of quantum nonlocality.

- The opposite to the above point also holds true: If two sources produce different density operators, ρ and σ, then it is ***always possible*** to find an operator A which leads to different probabilities when measured on ρ compared to when measured on σ, and hence can be used to distinguish between the two sources(In Problem Sheet 5 you will show this important fact).

- We introduced the density operator in order to describe within quantum theory a situation where a source emits states probabilistically. However, we can use the formalism of density operators to describe the old situation also – when a source emits the state $|\psi\rangle$. In this case, the density operator is simply
$$\rho = |\psi\rangle\langle\psi|, \tag{5.15}$$
i.e. the projector onto the state $|\psi\rangle$. It will prove useful to introduce a bit of terminology for this case: we say that this is a ***pure*** state. Conversely, when the source genuinely emits two or more states (meaning that the density operator is ***not*** a pure state, then we call it a ***mixed*** state.

- One advantage of using the density operator, even to describe pure states, is that it automatically takes care of the unphysical global phase of a quantum state. In particular, recall that both $|\psi\rangle$ and $|\psi'\rangle = e^{i\phi}|\psi\rangle$ correspond to the same physical state. At the level of density operators
$$\rho' = e^{i\phi}|\psi\rangle\langle\psi|e^{-i\phi} = |\psi\rangle\langle\psi| = \rho \tag{5.16}$$

- The other extreme type of density operator one often encounters is where we mix with uniform probability a complete basis of orthonormal states. For example, for a qubit, if the source emits $|0\rangle$ and $|1\rangle$, or $|+\rangle$ and $|-\rangle$ with equal probability. More generally, for a d level system, if the source emits one of d states $|e_k\rangle$ with probability $\frac{1}{d}$, where $|e_k\rangle$ form an arbitrary orthonormal basis of a Hilbert space of dimension d. In this case

$$\rho = \frac{1}{d}\sum_k |e_k\rangle\langle e_k| = \frac{1}{d}I \tag{5.17}$$

  where I is the identity operator, and we used the fact that any orthonormal set of vectors forms a resolution of the identity. This state is known as the ***maximally mixed state*** and is the density operator analogue of a uniform probability distribution.

Mathematically, the density operator also has a number of essential properties, all of which can be derived from the fundamental form $\rho = \sum_n p_n|\psi_n\rangle\langle\psi_n|$.

- ρ is ***Hermitian*** (or self-adjoint):

$$\rho = \rho^\dagger. \tag{5.18}$$

- ρ is a ***positive*** operator. This means on the one hand that all of the eigenvalues are non-negative. Often more useful, this is equivalent to $\langle u|\rho|u\rangle \geq 0$ for all $|u\rangle$, which we can easily verify:

$$\langle u|\rho|u\rangle = \sum_n p_n\langle u|\psi_n\rangle\langle\psi_n|u\rangle = \sum_n p_n|\langle\psi_n|u\rangle|^2 \geq 0. \tag{5.19}$$

  This encodes the fact that the density operator always leads to positive probabilities.

- ρ has ***unit trace***, which encodes both the normalisation of the quantum states, and probabilities:

$$\begin{aligned}\mathrm{tr}(\rho) &= \mathrm{tr}\left(\sum_n p_n|\psi_n\rangle\langle\psi_n|\right), \\ &= \sum_n p_n\langle\psi_n|\psi_n\rangle, \\ &= \sum_n p_n, \\ &= 1.\end{aligned} \tag{5.20}$$

- Any positive operator with unit trace corresponds to a valid density operator, and can be realised by at least one source. In particular, ρ can always be written in diagonal form, in terms of its eigenvalues and eigenvectors as

$$\rho = \sum_n \lambda_n|v_n\rangle\langle v_n|, \tag{5.21}$$

  such that $\lambda_n \geq 0$, $\sum_n \lambda_n = 1$ and $\langle v_m|v_n\rangle = \delta_{m,n}$. Thus, a source which emits the states $|v_n\rangle$ with probability λ_n will be described by such a density operator.

- As with any operator, it can sometimes be helpful to represent the density operator as a matrix in a particular basis. In this context, (and sometimes even more generally), the density operator is instead referred to as a ***density matrix***.

  As an example, for a qubit, the density operator ρ is represented as a matrix in the standard basis as

$$\rho = \begin{bmatrix} \langle 0|\rho|0\rangle & \langle 0|\rho|1\rangle \\ \langle 1|\rho|0\rangle & \langle 1|\rho|1\rangle \end{bmatrix}. \tag{5.22}$$

  This is particularly useful when finding the eigenvectors and eigenvalues of ρ.

5.3 Quantum theory with density operators

It is possible to reformulate all the axioms of quantum theory directly in terms of density operators, which can then be applied directly to both pure and mixed states.

- ***Axiom 1*** (states): The state of any physical system is represented by a density operator ρ, that is a positive operator on a Hilbert space, which satisfies the normalisation condition $\text{tr}(\rho) = 1$.
- ***Axiom 2*** (Evolution): Evolution (for a closed system) is described by a unitary operator U, such that
$$\rho \to \rho' = U\rho U^\dagger. \tag{5.23}$$
- ***Axiom 3*** (Measurement): Observables are associated with Hermitian operators. For the operator $A = \sum_n a_n P_n$,
  - The probability of obtaining the outcome a_n is
$$\text{Prob}(a_n) = \text{tr}(\rho P_n). \tag{5.24}$$
  - The expectation (or expected) value of A is
$$\langle A \rangle = \text{tr}(\rho A). \tag{5.25}$$
  - The state after obtaining the result a_n is
$$\rho' = \frac{P_n \rho P_n}{\text{tr}(\rho P_n)} \tag{5.26}$$

Example Let's return to the example from previously, where we measured $Z = |0\rangle\langle 0| - |1\rangle\langle 1|$ on the density operator $\rho = \frac{1}{4}|0\rangle\langle 0| + \frac{3}{4}|+\rangle\langle +|$. We saw previously that

$$\text{Prob}(+1) = \text{tr}(\rho|0\rangle\langle 0|) = \tfrac{5}{8}, \tag{5.27}$$

from which it follows that

$$\text{Prob}(-1) = 1 - \text{Prob}(+1) = \tfrac{3}{8}, \tag{5.28}$$

which can also be verified to be equal to $\text{tr}(\rho|1\rangle\langle 1|)$,

$$\begin{aligned} \text{Prob}(-1) &= \text{tr}\left(\left(\tfrac{1}{4}|0\rangle\langle 0| + \tfrac{3}{4}|+\rangle\langle +|\right)|1\rangle\langle 1|\right), \\ &= \text{tr}\left(\tfrac{1}{4}|0\rangle\langle 0|1\rangle\langle 1| + \tfrac{3}{4}|+\rangle\langle +|1\rangle\langle 1|\right), \\ &= \tfrac{1}{4}\langle 1|0\rangle\langle 0|1\rangle + \tfrac{3}{4}\langle +|1\rangle\langle 1|+\rangle, \\ &= \tfrac{3}{8}. \end{aligned} \tag{5.29}$$

We also see that

$$\begin{aligned} \langle Z \rangle &= \text{tr}\left(\rho(|0\rangle\langle 0| - |1\rangle\langle 1|)\right) \\ &= \text{tr}(\rho|0\rangle\langle 0|) - \text{tr}(\rho|1\rangle\langle 1|) \\ &= \text{Prob}(+1) - \text{Prob}(-1) \\ &= \tfrac{5}{8} - \tfrac{3}{8} \\ &= \tfrac{1}{4}. \end{aligned} \tag{5.30}$$

The state after the outcome $+1$ is given by

$$\begin{aligned}\rho' &= \frac{|0\rangle\langle 0|\rho|0\rangle\langle 0|}{\mathrm{tr}(\rho|0\rangle\langle 0|)}, \\ &= \frac{\langle 0|\rho|0\rangle|0\rangle\langle 0|}{\langle 0|\rho|0\rangle} \\ &= |0\rangle\langle 0|. \end{aligned} \tag{5.31}$$

That is, after a measurement of Z, if the outcome $+1$ is obtained, the system is left in the state $|0\rangle$ with certainty. This is true even if the state of the system was initially a mixed state.

We end this section by returning to the trace, and outlining a number of its important properties, all of which will prove useful in the remainder of this course.

5.4 Properties of the trace

The trace has a number of equivalent representations, each of which can prove useful.

- First, consider the trace of an arbitrary operator A, by inserting the identity operator $I = \sum_j |j\rangle\langle j|$:

$$\begin{aligned}\mathrm{tr}(A) &= \mathrm{tr}\left(A\sum_j |j\rangle\langle j|\right), \\ &= \sum_j \mathrm{tr}\left(A|j\rangle\langle j|\right), \\ &= \sum_j \langle j|A|j\rangle. \end{aligned} \tag{5.32}$$

  This is a useful form, and often the most direct way to calculate the trace. Informally, if we want to take the trace of an operator A (or any other expression, such as AB), then we can simply 'sandwich' the operator with $\langle i|$ and $|i\rangle$ and take the sum.

- If instead we consider the matrix of A in the basis $|i\rangle$, and recalling that the matrix elements of A in this basis are $A_{ij} = \langle i|A|j\rangle$,

$$\mathrm{tr}(A) = \sum_i \langle i|A|i\rangle = \sum_i A_{ii}. \tag{5.33}$$

  That is, the trace of an operator corresponds to the sum of the diagonal elements of the matrix in the basis. Thus, if we have an operator already in matrix form, we can again simply perform the trace by summing the diagonal entries.

- In the case that A is Hermitian, a natural choice of basis is the eigenbasis of A. If we denote the eigenvalues and eigenvectors by λ_i and $|e_i\rangle$ respectively, then

$$\mathrm{tr}(A) = \sum_i \langle e_i|A|e_i\rangle = \sum_i \lambda_i. \tag{5.34}$$

  That is, for Hermitian operators, the trace equals to the sum of the eigenvalues. This makes it particularly easy to take the trace of an operator if we know its eigenvalues.

- Finally, one useful property that the trace has is **cyclic symmetry**. In particular,

$$\begin{aligned}\mathrm{tr}(AB) &= \mathrm{tr}(BA) \quad \text{for all } A, B, \\ \mathrm{tr}(ABC) &= \mathrm{tr}(BCA) = \mathrm{tr}(CAB) \quad \text{for all } A, B, C. \end{aligned} \tag{5.35}$$

  These properties will be proved in the Problem Sheet 5.

6 *(Optional)* The Bloch Sphere

The density operators of single qubits have a particularly nice geometric representation, known as the ***Bloch sphere***.

Consider an arbitrary (pure) qubit state, which can always be written as

$$|\psi\rangle = e^{i\alpha}\left(\cos\left(\frac{\theta}{2}\right)|0\rangle + e^{i\phi}\sin\left(\frac{\theta}{2}\right)|1\rangle\right), \tag{6.1}$$

where we have chosen to write the amplitude of $|0\rangle$ as $e^{i\alpha}\cos(\theta/2)$ and the amplitude of $|1\rangle$ as $e^{i(\alpha+\phi)}\sin(\theta/2)$. This is without any loss of generality.

The density matrix of this state, in the standard basis is

$$|\psi\rangle\langle\psi| = \begin{pmatrix} \cos^2\left(\frac{\theta}{2}\right) & e^{-i\phi}\cos\left(\frac{\theta}{2}\right)\sin\left(\frac{\theta}{2}\right) \\ e^{i\phi}\cos\left(\frac{\theta}{2}\right)\sin\left(\frac{\theta}{2}\right) & \sin^2\left(\frac{\theta}{2}\right) \end{pmatrix}, \tag{6.2}$$

which is notably independent of α, which was a global phase.

Using the double angle formulas and Euler's formula

$$\cos\theta = 2\cos^2\left(\tfrac{\theta}{2}\right) - 1 = 1 - 2\sin^2\left(\tfrac{\theta}{2}\right), \tag{6.3}$$

$$\sin\theta = 2\sin\left(\tfrac{\theta}{2}\right)\cos\left(\tfrac{\theta}{2}\right), \tag{6.4}$$

$$e^{i\phi} = \cos(\phi) + i\sin(\phi) \tag{6.5}$$

we see that $|\psi\rangle\langle\psi|$ can be equivalently written as

$$\begin{aligned} |\psi\rangle\langle\psi| &= \frac{1}{2}\begin{pmatrix} 1+\cos\theta & (\cos\phi - i\sin\phi)\sin\theta \\ (\cos\phi + i\sin\phi)\sin\theta & 1-\cos\theta \end{pmatrix}, \\ &= \frac{1}{2}\left(\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \cos\phi\sin\theta\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} + \sin\phi\sin\theta\begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} + \cos\theta\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}\right), \\ &= \tfrac{1}{2}\left(I + (\sin\theta\cos\phi)X + (\sin\theta\sin\phi)Y + (\cos\theta)Z\right) \\ &= \tfrac{1}{2}(I + \hat{\mathbf{n}}\cdot\boldsymbol{\sigma}), \end{aligned} \tag{6.6}$$

where in the second line we wrote the matrix as a sum of matrices, which we realise are just the matrix representations of the Pauli operators, and in the final line we have defined the ***Bloch vector***

$$\hat{\mathbf{n}} = (\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta), \tag{6.7}$$

which is a ***unit vector*** in three dimensions, pointing in the direction θ, ϕ in spherical polar co-ordinates, and $\boldsymbol{\sigma} = (X, Y, Z)$ is a vector of Pauli operators.

This shows that the density operator of a pure qubit can be represented by the Bloch vector $\hat{\mathbf{n}}$. We say that $\hat{\mathbf{n}}$ is the ***Bloch vector representation of a qubit***.

Since $\hat{\mathbf{n}}$ is a *unit vector*, pure states of qubits lie on the surface of what is known as the ***Bloch sphere***. It is also important to realise that θ and ϕ were arbitrary – we could choose them to take any values we like, and each value produces one unit vector in $\mathbb{R}^3$. What this means is that the entire surface of the sphere corresponds to pure quantum states.

Arbitrary density operators ρ – including ***mixed*** states – can also be represented by Bloch vectors. In particular, consider the mixed state $\rho = \sum_k p_k|\psi_k\rangle\langle\psi_k|$, then since each pure state $|\psi_k\rangle\langle\psi_k|$ can be written as $|\psi_k\rangle\langle\psi_k| = \frac{1}{2}\left(I + \hat{\mathbf{n}}^{(k)}\cdot\boldsymbol{\sigma}\right)$ from above, for an appropriately defined Bloch vector

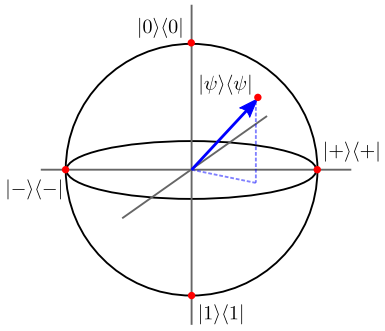

Figure 1: The Bloch sphere. Pure states lie on the surface, while mixed states lie in the interior. The centre of the sphere corresponds to the maximally mixed state $I/2$.

$\hat{\mathbf{n}}^{(k)}$, then it follows that

$$\begin{aligned}\rho &= \sum_k p_k |\psi_k\rangle\langle\psi_k|, \\ &= \sum_k p_k \tfrac{1}{2}\left(I + \hat{\mathbf{n}}^{(k)} \cdot \boldsymbol{\sigma}\right) \\ &= \frac{1}{2}\left(I + \left(\sum_k p_k \hat{\mathbf{n}}^{(k)}\right) \cdot \boldsymbol{\sigma}\right), \\ &= \tfrac{1}{2}\left(I + \mathbf{n} \cdot \boldsymbol{\sigma}\right). \end{aligned} \tag{6.8}$$

where in the final line we have defined an ***average Bloch vector***

$$\mathbf{n} = \sum_k p_k \hat{\mathbf{n}}^{(k)}, \tag{6.9}$$

which is the average of the Bloch vectors of the pure states in the decomposition of the density operator.

We should notice now that

$$\begin{aligned}\mathbf{n} \cdot \mathbf{n} &= \sum_{kl} p_k p_l \hat{\mathbf{n}}^{(k)} \cdot \hat{\mathbf{n}}^{(l)}, \\ &\leq \sum_{kl} p_k p_l, \\ &= 1. \end{aligned} \tag{6.10}$$

That is, the Bloch vector $\mathbf{n}$ of a density operator ρ is never longer than a unit vector, but in general can be shorter. That is, these vectors are in general inside the unit sphere.

It can also be shown that equality will never be reached if the state is mixed (this is left as an exercise in the Problem Sheets). We can conclude therefore that ***density operators corresponding to mixed states are in the interior of the sphere***, which contrasts to *pure* states which live on the surface. It can also be shown (but will not be shown here), that ***any*** vector $\mathbf{n}$ inside the Bloch sphere, corresponds to a valid density operator. That is, ***the space of qubits is in one-to-one correspondence with the set of Bloch vectors***, which is given by the unit sphere in $\mathbb{R}^3$. This gives us a very nice and intuitive way to visualise qubits.

An interesting limiting case is that of the maximally mixed state

$$\rho = \tfrac{1}{2}I = \tfrac{1}{2}|0\rangle\langle 0| + \tfrac{1}{2}|1\rangle\langle 1|. \tag{6.11}$$

The Bloch vector is $\mathbf{n} = (0,0,0)$, i.e. the centre of the Bloch sphere.

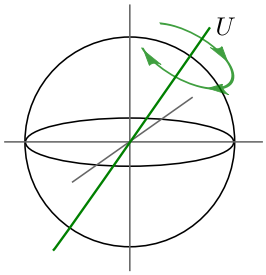

Figure 2: Unitary tranformations correspond to rotations of the Bloch sphere.

There is a quick method in order to calculate the Bloch vector from a given density operator. The components are given by

$$n_x = \mathrm{tr}(X\rho), \qquad n_y = \mathrm{tr}(Y\rho) \qquad n_z, = \mathrm{tr}(Z\rho), \tag{6.12}$$

This follows by direct calculation:

$$\begin{aligned} \mathrm{tr}(X\rho) &= \mathrm{tr}\left(X\tfrac{1}{2}\left(I + \mathbf{n}\cdot\boldsymbol{\sigma}\right)\right), \\ &= \mathrm{tr}\left(X\tfrac{1}{2}\left(I + n_x X + n_y Y + n_z Z\right)\right), \\ &= \tfrac{1}{2}\left(\mathrm{tr}(X) + n_x\,\mathrm{tr}(X^2) + n_y\,\mathrm{tr}(XY) + n_z\,\mathrm{tr}(XZ)\right), \\ &= n_x. \end{aligned} \tag{6.13}$$

For n_y and n_z the analogous calculations follow. Thus, the components of the Bloch vector are precisely the expectation values of the three Pauli spin observables in the state.

One surprising and counter-intutive feature of the Bloch sphere representation of qubits is that ***orthogonal pure quantum states correspond to opposite points on the surface of the Bloch sphere!*** That is, if a pure quantum state $|\psi\rangle$ has Bloch vector $\hat{\mathbf{n}}$, then the orthogonal state $|\psi^\perp\rangle$, such that $\langle\psi^\perp|\psi\rangle = 0$, has Bloch vector $\hat{\mathbf{n}}^\perp = -\hat{\mathbf{n}}$. Therefore we see that

$$\hat{\mathbf{n}}^\perp \cdot \hat{\mathbf{n}} = -1. \tag{6.14}$$

Thus, ***orthogonality of the quantum states is not the same as orthogonality of the Bloch vectors*** – an important cautionary point.

Finally, if we consider applying a unitary transformation (to an arbitrary qubit), we can ask what happens to the Bloch sphere. That is, we can think ***globally*** about the action of a unitary transformation on the entire space of quantum states. An appealing feature of the Bloch sphere representation is that unitary transformations correspond to ***rotations*** of the Bloch sphere.

Similar, if we consider what a measurement looks like for the Bloch sphere, we find that it depends upon the projection of the Bloch vector along an axis (defined by the measurement). For example, when measuring Z on a qubit, the probability to obtain $+1$ is

$$\mathrm{Prob}(+1) = \mathrm{tr}(\rho|0\rangle\langle 0|) = \frac{1}{2}(1 + n_z). \tag{6.15}$$

7 Reduced Density Operators

7.1 The reduced density operator

Suppose that Alice and Bob each have a particle, and their systems are in a ***pure product state***, e.g.

$$|\chi\rangle = |0\rangle|+\rangle. \tag{7.1}$$

In this case, it is rather straightforward to answer the question "*what is the state of the first system?*". The answer is $|0\rangle$, and indeed, if we were to consider ***ignoring*** the second system, we would use $|0\rangle$ to predict the outcome of any measurement made solely on the first system.

But what if we consider not a product state, but an ***entangled state***, e.g. consider that Alice and Bob share the Bell state

$$|\Phi\rangle = \frac{|0\rangle|0\rangle + |1\rangle|1\rangle}{\sqrt{2}}. \tag{7.2}$$

Now it is not so easy to see what the state of the first system is. Indeed, the fact that the system is entangled precisely means that we cannot factor it into one state for each system. In order to try and understand what the answer should be, let's consider ignoring Bob's system, and ask Alice to make measurements of the Pauli operators X, Y and Z on her qubit. If we were to calculate, we would find that for all 3 measurements, the probability to obtain $+1$ and -1 are equal. However, there is no pure qubit state $|\psi\rangle$ which has this property!

We are going to show below that Alice's qubit, when considered alone, behaves as if they were in the ***maximally mixed state*** $\rho = \frac{1}{2}I = \frac{1}{2}(|0\rangle\langle 0| + |1\rangle\langle 1|)$. That is, her qubit is described by a ***density operator*** corresponding to a mixed state, and not a pure quantum state!

We will in fact see that this is a hallmark of entanglement – when particles are entangled, subsystems necessarily become mixed quantum states, described by density operators. This is in fact one of the primary reasons why we need to introduce density operators, to allow us to consistently describe ***parts*** of entangled states.

Note that here, in contrast to Section 5, where the density operator was introduced to account for *ignorance* (i.e. we had a source, and ***we*** didn't know which state was produced, only the probabilities), here there is no ignorance whatsoever. Nevertheless, remarkably we still need to use the density operator to correctly describe the subsystem.

We will now show how this result arises. Suppose that Alice and Bob share the maximally entangled state $|\Phi\rangle$, and Alice measures $X = |+\rangle\langle +| - |-\rangle\langle -|$ on her subsystem. As a measurement on the bipartite system this operator is $X \otimes I$. Written out in spectral decomposition, this is

$$X \otimes I = |+\rangle\langle +| \otimes I - |-\rangle\langle -| \otimes I. \tag{7.3}$$

The probability for Alice to obtain the result $+1$ is

$$\begin{aligned}
\text{Prob}(+1) &= \left\| (|+\rangle\langle +| \otimes I) \frac{|0\rangle|0\rangle + |1\rangle|1\rangle}{\sqrt{2}} \right\|^2, \\
&= \left\| \frac{\langle +|0\rangle|+\rangle|0\rangle + \langle +|1\rangle|+\rangle|1\rangle}{\sqrt{2}} \right\|^2, \\
&= \left\| |+\rangle \frac{\langle +|0\rangle|0\rangle + \langle +|1\rangle|1\rangle}{\sqrt{2}} \right\|^2, \\
&= \frac{\langle +|0\rangle\langle 0|+\rangle + \langle +|1\rangle\langle 1|+\rangle}{2}, \\
&= \langle +| \left(\frac{|0\rangle\langle 0| + |1\rangle\langle 1|}{2} \right) |+\rangle,
\end{aligned} \tag{7.4}$$

where in the second line we carefully applied the projector $|+\rangle\langle+| \otimes I$ onto $|\Phi\rangle$, in the third line we factored the state (since in both terms the first qubit is in the state $|+\rangle$).

The operator inside the brackets is the ***maximally mixed state*** $I/2$, and our formula says that $Prob(+1) = \langle+|I/2|+\rangle$. This is exactly of the form (5.7) – it is the formula we would use if wanted to calculate the probability of obtaining the result $+1$ when measuring X on the maximally mixed state $I/2$. This indicates that as far as a measurement of X is concerned, Alice's qubit is behaving identical to the maximally mixed state $I/2$.

We could carry out exactly the same calculation for a measurement of Y or Z – or ***any other operator*** – and we would arrive at the same conclusion, that the statistics of any measurement would be identical to the statistics from the maximally mixed state. For this reason, ***we conclude that this is the correct description of Alice's qubit*** – that her subsystem is the maximally mixed state.

We call this her ***reduced density operator***. It is the density operator that describes her subsystem (if we are interested ***only*** in this subsystem). We denote this density operator by ρ_A, where the ***subscript*** A signifies that this is a reduced density operator (i.e. describing a subsystem), and that this is the subsystem of Alice. If we instead were only interested in Bob's subsystem, then it too is described by a reduced density operator, which we would then denote by ρ_B.

In general, how do we find the reduced density operator for an arbitrary state $|\Psi\rangle$? Luckily there are two very nice ways to do so. The first by ***extending the definition of the trace*** to subsystems, called the ***partial trace***. The second way – which is in fact equivalent – is a fast method for carrying out the partial trace.

The partial trace

The ***partial trace*** is a way of applying the trace operation to only one part of an operator that acts on a tensor product space, rather than to the whole operator.

The ***partial trace*** **over the second space** is the *linear* operation that performs the trace on the second space, but leaves the first unchanged,

$$\mathrm{tr}_2\left(|u\rangle\langle v| \otimes |w\rangle\langle x|\right) = \langle x|w\rangle|u\rangle\langle v| \quad \text{for all } |u\rangle, |v\rangle, |w\rangle, |x\rangle, \tag{7.5}$$

which should be compared with Eq. (5.8). The partial trace over the second system takes operators that act jointly on both Hilbert spaces to operators that act only on the first.

Colloquially the partial trace is referred to as ***tracing out*** a space, as it 'removes' the space that it acts on, by converting it to a complex number.

Analogously, the operation of ***tracing out the first space***, or more formally ***the partial trace over the first space*** is the linear operation that satisfies

$$\mathrm{tr}_1\left(|u\rangle\langle v| \otimes |w\rangle\langle x|\right) = \langle v|u\rangle|w\rangle\langle x| \quad \text{for all } |u\rangle, |v\rangle, |w\rangle, |x\rangle, \tag{7.6}$$

i.e. we have simply applied the trace on the first Hilbert space instead of the second Hilbert space.

Sometimes it is natural to label the two Hilbert spaces as A and B, rather than just referring to 'the first' or 'the second', (when they are associated to Alice and Bob, respectively). In this case, we write tr_A instead of tr_1, and refer colloquially to ***tracing out A*** or ***tracing out Alice***. That is, the subscript on the partial trace is a ***label***, and what is important is that the label is meaningful and unambiguous.

Let's apply this to the maximally entangled state $|\Phi\rangle$ and confirm that we get back $I/2$ as the

reduced density operator for Alice, as we found above:

$$\begin{aligned}\mathrm{tr}_B(|\Phi\rangle\langle\Phi|) &= \mathrm{tr}_B\left(\left(\frac{|0\rangle|0\rangle+|1\rangle|1\rangle}{\sqrt{2}}\right)\left(\frac{\langle 0|\langle 0|+\langle 1|\langle 1|}{\sqrt{2}}\right)\right),\\ &= \frac{1}{2}\,\mathrm{tr}_B\left(|0\rangle\langle 0|\otimes|0\rangle\langle 0|+|0\rangle\langle 1|\otimes|0\rangle\langle 1|+|1\rangle\langle 0|\otimes|1\rangle\langle 0|+|1\rangle\langle 1|\otimes|1\rangle\langle 1|\right),\\ &= \frac{1}{2}\left(\langle 0|0\rangle|0\rangle\langle 0|+\langle 1|0\rangle|0\rangle\langle 1|+\langle 0|1\rangle|1\rangle\langle 0|+\langle 1|1\rangle|1\rangle\langle 1|\right),\\ &= \frac{1}{2}|0\rangle\langle 0|+\frac{1}{2}|1\rangle\langle 1| = \frac{1}{2}I. \end{aligned} \tag{7.7}$$

Note that in the second line we used the formula (3.24), that $|i\rangle|j\rangle\langle k|\langle\ell| \equiv |i\rangle\langle k|\otimes|j\rangle\langle\ell|$, which is nothing but ***two different convenient notations for the same operator***.

Let us now see how to carry out the partial trace more indirectly, and effectively 'write down' the reduced density operator directly. Let's consider an arbitrary pure state of two qubits,

$$|\Psi\rangle = \alpha|0\rangle|0\rangle+\beta|0\rangle|1\rangle+\gamma|1\rangle|0\rangle+\delta|1\rangle|1\rangle. \tag{7.8}$$

On the one hand, we can just calculate the reduced density operator (of the first subsystem) directly, namely

$$\begin{aligned}\rho_1 &= \mathrm{tr}_2\left(|\Psi\rangle\langle\Psi|\right),\\ &= \mathrm{tr}_2\left((\alpha|0\rangle|0\rangle+\beta|0\rangle|1\rangle+\gamma|1\rangle|0\rangle+\delta|1\rangle|1\rangle)\,\alpha^*\langle 0|\langle 0|+\beta^*\langle 0|\langle 1|+\gamma^*\langle 1|\langle 0|+\delta^*\langle 1|\langle 1|\right),\\ &= \mathrm{tr}_2\left(|\alpha|^2|0\rangle\langle 0|\otimes|0\rangle\langle 0|+\alpha\beta^*|0\rangle\langle 0|\otimes|0\rangle\langle 1|+\alpha\gamma^*|0\rangle\langle 1|\otimes|0\rangle\langle 0|+\alpha\delta^*|0\rangle\langle 1|\otimes|0\rangle\langle 1|+\ldots\right),\\ &= |\alpha|^2|0\rangle\langle 0|\langle 0|0\rangle+\alpha\beta^*|0\rangle\langle 0|\langle 1|0\rangle+\alpha\gamma^*|0\rangle\langle 1|\langle 0|0\rangle+\alpha\delta^*|0\rangle\langle 1|\langle 1|0\rangle+\ldots,\\ &= |\alpha|^2|0\rangle\langle 0|+\alpha\gamma^*|0\rangle\langle 1|+\alpha^*\gamma|1\rangle\langle 0|+|\gamma|^2|1\rangle\langle 1|\\ &\qquad+|\beta|^2|0\rangle\langle 0|+\beta\delta^*|0\rangle\langle 1|+\beta^*\delta|1\rangle\langle 0|+|\delta|^2|1\rangle\langle 1|, \end{aligned} \tag{7.9}$$

where in the second and third line we have omitted (twelve) terms for brevity. What we see, when we arrive at the fourth line is that ***many terms cancel*** when performing the partial trace, in particular ***all those terms where the ket and the bra differ***. This means that if we take this approach of applying the partial trace, we will always end up writing down many terms that are going to cancel. The basic idea behind this method is to ***identify which terms are going to cancel, and avoid ever writing them down in the first place!*** In order to do this, if we look at the final equality above, we notice that we can actually factorise the final 8 terms, namely

$$\rho_1 = (\alpha|0\rangle+\gamma|1\rangle)(\alpha^*\langle 0|+\gamma^*\langle 1|)+(\beta|0\rangle+\delta|1\rangle)(\beta^*\langle 0|+\delta^*\langle 1|). \tag{7.10}$$

On the other hand, if we look back at the state (7.8), we notice that we can also factorise it similarly,

$$|\Psi\rangle = (\alpha|0\rangle+\gamma|1\rangle)|0\rangle+(\beta|0\rangle+\delta|1\rangle)|1\rangle, \tag{7.11}$$

in terms of ***basis states in the system being traced out*** (in this case, in terms of the second qubit), with exactly the same two states which ρ_1 factorised into. This is completely general, and provides us with our method for ***writing down*** a reduced density operator: we first factorise it, and then uses these states to write down the density operator.

As an example, let's use this method to write down ρ_2 $(= \mathrm{tr}_1(|\Psi\rangle\langle\Psi|)$ directly. Since we want to trace out the first qubit, we start by writing down $|\Psi\rangle$ factorised now in terms of this qubit,

$$|\Psi\rangle = |0\rangle(\alpha|0\rangle+\beta|1\rangle)+|1\rangle(\gamma|0\rangle+\delta|1\rangle). \tag{7.12}$$

It is then the states $\alpha|0\rangle+\beta|1\rangle$ and $\gamma|0\rangle+\delta|1\rangle$ of the second qubit, which appear alongside $|0\rangle$ and $|1\rangle$ on the first system respectively, which form the density operator of qubit 2, namely

$$\rho_2 = (\alpha|0\rangle+\beta|1\rangle)(\alpha^*\langle 0|+\beta^*\langle 1|)+(\gamma|0\rangle+\delta|1\rangle)(\gamma^*\langle 0|+\delta^*\langle 1|) \tag{7.13}$$

As a final calculation, let's return to the maximally entangled state $|\Phi\rangle = \frac{1}{\sqrt{2}}|0\rangle|0\rangle + \frac{1}{\sqrt{2}}|1\rangle|1\rangle$ to double check that we obtain the expected result for its reduced density operator. In this case, we see that the state is factorised as written, and hence applying this method we immediately write down

$$\begin{aligned}\rho_A &= \left(\frac{1}{\sqrt{2}}|0\rangle\right)\left(\frac{1}{\sqrt{2}}\langle 0|\right) + \left(\frac{1}{\sqrt{2}}|1\rangle\right)\left(\frac{1}{\sqrt{2}}\langle 1|\right), \\ &= \frac{1}{2}|0\rangle\langle 0| + \frac{1}{2}|1\rangle\langle 1|, \end{aligned} \tag{7.14}$$

as we had already calculated using the partial trace. With a little practice, this method makes it very fast to quickly write down the reduced density operator in many cases of interest.

As a final word of caution, the reduced density operator provides a complete description of a ***subsystem alone***. It is very useful when we want to consider properties of a subsytem in isolation, as it throws away all superfluous information. If however we want to study any property not about the subsystem – e.g. ***correlations*** between two subsystems, ***then we definitely cannot use the reduced density operator!*** In this case, we must use the full state $|\Psi\rangle$, as we would have done before learning about the reduced density operator.

7.2 Identifying entanglement using the reduced density operator

In the above we saw that when we asked the question 'what is the state of a subsystem?', when the state is entangled, the correct answer is a (mixed) density operator. Here we will see that this is always true – and in fact provides us with a nice method for identifying whether a state is entangled or not. In particular, the following important result holds

> $|\Psi\rangle$ is entangled if and only if ρ_A is a ***mixed*** density operator,

where we recall that being ***mixed*** means that the density operator is a mixture of at least two pure states (which is the opposite of being ***pure***).

It is easy to check that if the state is *not* entangled – i.e. is a product state – then the reduced density operator is a ***pure*** state, and therefore *not* mixed. In particular, if the state is

$$|\Psi\rangle = |\psi\rangle|\phi\rangle, \tag{7.15}$$

which is an arbtitrary product state, then the reduced density operator for Alice is

$$\rho_A = \mathrm{tr}_B(|\psi\rangle\langle\psi| \otimes |\phi\rangle\langle\phi|) = |\psi\rangle\langle\psi|, \tag{7.16}$$

which is a pure state.

A proof of the other direction, that if the state is entangled then the reduced density operator is necessarily mixed, is not as straightforward to demonstrate, and is beyond the scope of this course. We will however use this result without proof.

We can go one step further, and realise that it is easy to check whether a reduced density operator is mixed or not. In particular, pure density operators correspond to projectors of the form $P = |\chi\rangle\langle\chi|$ for some state $|\chi\rangle$. As projectors, they satisfy $P^2 = P$, which can easily be confirmed

$$P^2 = (|\chi\rangle\langle\chi|)(|\chi\rangle\langle\chi|) = |\chi\rangle\langle\chi|\chi\rangle\langle\chi| = |\chi\rangle\langle\chi| = P \tag{7.17}$$

Thus, when a density operator ρ_A corresponds to a pure state, we have the relation $\rho_A^2 = \rho_A$. On the other hand, in the problem sheets you will show that ***only*** pure density operators satisfy this relation. That is, if the reduced density operator is a mixed state, then we have $\rho_A^2 \neq \rho_A$. What this shows is that the ***mixedness*** of a density operator can be determined by simply comparing ρ^2 with ρ. We thus arrive at a very useful method for detecting entanglement

$|\Psi\rangle$ is entangled if and only if $\rho_A^2 \neq \rho_A$,

where $\rho_A = \mathrm{tr}_B(|\Psi\rangle\langle\Psi|)$.
We could have alternatively stated all of the above in terms of ρ_B instead, and we would arrive at the same conclusion. For simplicity, we state it here only in terms of ρ_A (since it is neither easier nor harder to find ρ_A than ρ_B).

Example Consider the state $|\Phi\rangle = \sqrt{\frac{2}{3}}|0\rangle|0\rangle + \sqrt{\frac{1}{3}}|1\rangle|1\rangle$, then

$$\rho_A = \mathrm{tr}_B(|\Phi\rangle\langle\Phi|) = \tfrac{2}{3}|0\rangle\langle 0| + \tfrac{1}{3}|1\rangle\langle 1|, \tag{7.18}$$

and

$$\rho_A^2 = \tfrac{4}{9}|0\rangle\langle 0| + \tfrac{1}{9}|1\rangle\langle 1| \tag{7.19}$$

which is not equal to ρ_A, and hence $|\Phi\rangle$ is entangled.

8 *(Optional)* No signalling via collapse

In this section we will show that quantum mechanics has a special property, which we refer to as being ***non-signalling***, which means that it is consistent with relativity. In particular, from relativity we know that it should not be possible that actions in one place lead instantaneously to observable effects in distant places, as this type of ***signalling*** leads to paradoxes such as being able to time-travel. We will see that quantum mechanics satisfies this basic principle. It is important to show this, as it sets the scene for studying quantum nonlocality in the next section.

Let's consider the following seemingly problematic feature of quantum mechanics: when we perform a measurement on a quantum system, straight after the measurement the state of the system is projected – or more colloquially ***collapsed*** – onto the eigenstate corresponding to the observed outcome. Crucially, this remains true even if the measurement is performed ***locally*** on one part of a system which is entangled with another, distant, part.

This local measurement collapses the state of the ***whole***, i.e. a measurement by Alice leads to a collapse of the state of both Alice and Bob. There is thus seemingly action at a distance in quantum mechanics, and this should seemingly lead to signalling and paradoxes.

In what follows, we will see that this is not the case and that in fact quantum mechanics is non-signalling. Quantum mechanics escapes from being signalling precisely because a single density operator can arise in a number of different ways. It will turn out that this property is precisely what is needed in order to avoid any observable action at a distance.

We will first study in detail the case of the maximally entangled state of two qubits, which contains all of the essential ingredients, before giving the general argument.

Suppose Alice and Bob share the maximally entangled state

$$|\Phi\rangle = \frac{1}{\sqrt{2}}(|0\rangle|0\rangle + |1\rangle|1\rangle). \tag{8.1}$$

Let us assume that Alice measures locally in her laboratory the observable $Z = |0\rangle\langle 0| - |1\rangle\langle 1|$. As an operator on the whole state this is $Z \otimes I$. The probability for her to obtain the result $+1$ is

$$\begin{aligned} \text{Prob}(+1) &= \|(|0\rangle\langle 0| \otimes I)|\Phi\rangle\|^2, \\ &= \left\| (|0\rangle\langle 0| \otimes I) \left(\frac{|0\rangle|0\rangle + |1\rangle|1\rangle}{\sqrt{2}} \right) \right\|^2 \\ &= \left\| \frac{|0\rangle|0\rangle}{\sqrt{2}} \right\|^2 \\ &= \frac{1}{2}. \end{aligned} \tag{8.2}$$

Let us denote the state right after the measurement by $|\Psi_+\rangle$. We see that it is

$$\begin{aligned} |\Psi_+\rangle &= \frac{(|0\rangle\langle 0| \otimes I)|\Phi\rangle}{\|(|0\rangle\langle 0| \otimes I)|\Phi\rangle\|}, \\ &= \frac{(|0\rangle\langle 0| \otimes I)\left(\frac{|0\rangle|0\rangle+|1\rangle|1\rangle}{\sqrt{2}}\right)}{\frac{1}{\sqrt{2}}} \\ &= |0\rangle|0\rangle. \end{aligned} \tag{8.3}$$

This tells us that right after Alice obtains the outcome $+1$, Bob is left in the state $|0\rangle$.

Similarly, the probability for Alice to obtain the result -1 is

$$\begin{aligned}\text{Prob}(-1) &= \|(|1\rangle\langle 1| \otimes I)|\Phi\rangle\|^2, \\ &= \left\|(|1\rangle\langle 1| \otimes I)\left(\frac{|0\rangle|0\rangle + |1\rangle|1\rangle}{\sqrt{2}}\right)\right\|^2 \\ &= \left\|\frac{|1\rangle|1\rangle}{\sqrt{2}}\right\|^2 \\ &= \frac{1}{2}. \end{aligned} \tag{8.4}$$

and the state immediately after obtaining this result, which we denote by $|\Psi_-\rangle$ is

$$\begin{aligned}|\Psi_-\rangle &= \frac{(|1\rangle\langle 1| \otimes I)|\Phi\rangle}{\|(|1\rangle\langle 1| \otimes I)|\Phi\rangle\|}, \\ &= \frac{(|1\rangle\langle 1| \otimes I)\left(\frac{|0\rangle|0\rangle+|1\rangle|1\rangle}{\sqrt{2}}\right)}{\frac{1}{\sqrt{2}}} \\ &= |1\rangle|1\rangle. \end{aligned} \tag{8.5}$$

Now, after Alice obtains the outcome -1, we see that Bob is left in the state $|1\rangle$.

Altogether, Alice's measurement of Z causes Bob's particle to collapse to either the state $|0\rangle$ or $|1\rangle$, with equal probability, depending on the result of Alice's measurement.

Crucially, since it was Alice who performed the measurement, the outcome is not be accessible or known to Bob straight away. Indeed, we are precisely interested in cases where Bob is ***separated*** from Alice, and as such, it is only Alice who knows the outcome of the measurement, and not Bob.

Therefore, let us think a little more carefully about the situation from Bob's perspective. All he knows are the probabilities with which the outcome of Alice's measurement will occur – in this instance $\text{Prob}(+1) = 1/2$ and $\text{Prob}(-1) = 1/2$. He will however be ***ignorant*** about the specific outcome Alice observed.

From his perspective, the situation is thus identical to the situation from Section 5: Bob can treat Alice like a source that either emits the state $|0\rangle$ or the state $|1\rangle$, both with equal probability. Here, Bob does not obtain the state directly from the source like previously, but his particle is collapsed into these states by the act of Alice's measurement.

Nevertheless, Bob should describe his system by the density operator corresponding to this source, i.e. by

$$\rho = \tfrac{1}{2}|0\rangle\langle 0| + \tfrac{1}{2}|1\rangle\langle 1|. \tag{8.6}$$

Let us now consider an alternative situation where Alice measures $X = |+\rangle\langle +| - |-\rangle\langle -|$ instead of Z.

Since $|0\rangle = \frac{|+\rangle+|-\rangle}{\sqrt{2}}$, and $|-\rangle = \frac{|0\rangle-|1\rangle}{\sqrt{2}}$, we see that, expressed in the basis of $|+\rangle$ and $|-\rangle$,

$$\begin{aligned}|\Phi\rangle &= \frac{1}{\sqrt{2}}\left(\left(\frac{|+\rangle + |-\rangle}{\sqrt{2}}\right)\left(\frac{|+\rangle + |-\rangle}{\sqrt{2}}\right) + \left(\frac{|+\rangle - |-\rangle}{\sqrt{2}}\right)\left(\frac{|+\rangle - |-\rangle}{\sqrt{2}}\right)\right) \\ &= \frac{1}{\sqrt{2}}\left(|+\rangle|+\rangle + |-\rangle|-\rangle\right). \end{aligned} \tag{8.7}$$

That is, the state $|\Phi\rangle$ has the same ***form*** when written in the basis of $|0\rangle$ and $|1\rangle$ as it does in the basis of $|+\rangle$ and $|-\rangle$.

Moreover, since X has the same form in the basis of $|+\rangle$ and $|-\rangle$ as Z does in the basis of $|0\rangle$ and $|1\rangle$, in this alternative situation everything must happen just as in the previous situation, but with $|+\rangle$ replacing $|0\rangle$ and $|-\rangle$ replacing $|1\rangle$.

That is, we again have that $\text{Prob}(+1) = \text{Prob}(-1) = \frac{1}{2}$. Now, after obtaining each outcome the state of the pair of systems will be either $|+\rangle|+\rangle$ or $|-\rangle|-\rangle$, respectively.

So now, Alice's measurement of X causes Bob's particle to collapse to either $|+\rangle$ or $|-\rangle$, with equal probability.

Just as above, since the outcome of Alice's measurement isn't known or accessible to Bob, he is ignorant of the actual outcome received, and from his perspective there is a source which is producing $|+\rangle$ and $|-\rangle$ with equal probability, and hence his state is the density operator

$$\rho' = \tfrac{1}{2}|+\rangle\langle+| + \tfrac{1}{2}|-\rangle\langle-|. \tag{8.8}$$

As we already saw in Section 5 this density operator is in fact equal to the density operator ρ, and both are equal to the maximally mixed state,

$$\rho = \rho' = \tfrac{1}{2}I. \tag{8.9}$$

This shows that, irrespective of whether Alice measured X or Z, even though the state of the pair of particles changes instantaneously – including the state of Bob's particle – because of the fact that Alice's outcome cannot be predicted in advance, and is not known to Bob, his particle is described by ***the same density operator*** in both cases.

This means, in particular, that if Bob doesn't know what Alice measured, there is absolutely no way that he can find out. Indeed, all probabilities of any measurement he can perform locally in his laboratory depend only upon his density operator, and they are the same in both cases. We thus conclude that ***there is no observable action at a distance when Alice collapses the state of Bob at a distance.***

We can also notice one further important point. The density operator that arises after either measurement also coincides with the reduced density operator of Bob

$$\begin{aligned}\rho_B &= \text{tr}_A(|\Phi\rangle\langle\Phi|) \\ &= \frac{1}{2}|0\rangle\langle 0| + |1\rangle\langle 1| = \frac{1}{2}I.\end{aligned} \tag{8.10}$$

This is important, as it shows that not only can Bob not tell which measurement Alice made, he furthermore cannot even tell if she made a measurement at all. ***Bob, locally, cannot tell whether or not Alice caused the state to collapse or not***.

We thus see that the instantaneous collapse of a quantum state after measurement does not allow for any form of signalling from Alice to Bob. This happens precisely because of the fact that the average state of Bob after different measurements of Alice can coincide (and coincide with the reduced density operator of Bob) which therefore mean that no observable effects of the collapse arise.

We can also view this as providing an answer to a puzzling question: Why should it be that the reduced density operator of an entangled state is a mixed state, when the joint state is pure? We can see that the reason is that otherwise we would have had a conflict with relativity, as only mixed density operators can be realised in a number of different ways (i.e. by a number of genuinely different sources).

8.1 The general argument

The above holds more generally:

> ***The collapsed state prepared for Bob by a measurement of Alice, averaged over the possible outcomes, is always equal to the reduced density operator of Bob.***

To see this, let us consider that Alice and Bob share the state $|\Psi\rangle$,

$$|\Psi\rangle = \sum_{ij} \alpha_{ij} |i\rangle|j\rangle, \tag{8.11}$$

which has reduced density operator for Bob

$$\begin{aligned}
\rho_B = \mathrm{tr}_A(|\Psi\rangle\langle\Psi|) &= \mathrm{tr}_A\left(\left(\sum_{ij}\alpha_{ij}|i\rangle|j\rangle\right)\left(\sum_{kl}\alpha_{kl}^*\langle k|\langle l|\right)\right), \\
&= \mathrm{tr}_A\left(\sum_{ijkl}\alpha_{ij}\alpha_{kl}^*|i\rangle\langle k|\otimes|j\rangle\langle l|\right), \\
&= \sum_{ijkl}\alpha_{ij}\alpha_{kl}^*\langle k|i\rangle|j\rangle\langle l|, \\
&= \sum_{ijl}\alpha_{ij}\alpha_{il}^*|j\rangle\langle l|.
\end{aligned} \tag{8.12}$$

Let us assume that Alice measures an arbitrary operator

$$C = \sum_m \lambda_m |e_m\rangle\langle e_m|, \tag{8.13}$$

which on the combined system is given by $C \otimes I = \sum_m \lambda_m (|e_m\rangle\langle e_m| \otimes I)$. The probability to obtain the outcome λ_m is then

$$\mathrm{Prob}(\lambda_m) = \|(|e_m\rangle\langle e_m| \otimes I)|\Psi\rangle)\|^2 \tag{8.14}$$

and the state immediately after the measurement is

$$\frac{(|e_m\rangle\langle e_m| \otimes I)|\Psi\rangle}{\|(|e_m\rangle\langle e_m| \otimes I)|\Psi\rangle)\|} = \frac{1}{\sqrt{\mathrm{Prob}(\lambda_m)}}|e_m\rangle\sum_{ij}\alpha_{ij}\langle e_m|i\rangle|j\rangle, \tag{8.15}$$

where we used the fact that $\|(|e_m\rangle\langle e_m| \otimes I)|\Psi\rangle)\| = \sqrt{\mathrm{Prob}(\lambda_m)}$, to re-express the denominator, which will be useful later.

The state that Bob is collapsed into can thus be read off, and is given by

$$|\psi_m\rangle = \frac{1}{\sqrt{\mathrm{Prob}(\lambda_m)}}\sum_{ij}\alpha_{ij}\langle e_m|i\rangle|j\rangle. \tag{8.16}$$

As in the above, since Bob does not know the outcome of the measurement, his final density operator will correspond to that which would arise from the (fictitious) source that emits the states $|\psi_m\rangle$ with probabilities $\mathrm{Prob}(\lambda_m)$. In particular, we find

$$\begin{aligned}
\rho &= \sum_m \mathrm{Prob}(\lambda_m)|\psi_m\rangle\langle\psi_m|, \\
&= \sum_m \mathrm{Prob}(\lambda_m)\left(\frac{1}{\sqrt{\mathrm{Prob}(\lambda_m)}}\sum_{ij}\alpha_{ij}\langle e_m|i\rangle|j\rangle\right)\left(\frac{1}{\sqrt{\mathrm{Prob}(\lambda_m)}}\sum_{kl}\alpha_{kl}^*\langle k|e_m\rangle\langle l|\right), \\
&= \sum_{ijklm}\alpha_{ij}\alpha_{kl}^*\langle k|e_m\rangle\langle e_m|i\rangle|j\rangle\langle l|, \\
&= \sum_{ijkl}\alpha_{ij}\alpha_{kl}^*\left[\langle k|\left(\sum_m|e_m\rangle\langle e_m|\right)|i\rangle\right]|j\rangle\langle l|, \\
&= \sum_{ijkl}\alpha_{ij}\alpha_{kl}^*\langle k|i\rangle|j\rangle\langle l|, \\
&= \sum_{ijl}\alpha_{ij}\alpha_{il}^*|j\rangle\langle l|,
\end{aligned} \tag{8.17}$$

where to go from the fourth line to the fifth line, we used the fact that $\sum_m |e_m\rangle\langle e_m| = I$ is a representation of the identity operator.

We see that this final post-measurement density operator of Bob is equal to the reduced density operator ρ_B as given in (8.12). Since the operator C measured was arbitrary, we see that no matter what Alice measures, quantum mechanics has the property that the density operator of Bob after the measurement is equal to his reduced density operator, and hence Bob has no way to determine what Alice measured, or whether or not she measured at all.

In this way, we see that the instantaneous collapse of the state after measurement, which is something that happens globally, nevertheless respects relativity and doesn't lead to instantaneous observable action at a distance.

This shouldn't however lead you to conclude that there is nothing strange about collapse in quantum mechanics. As we shall see in the next Section, collapse is a nonlocal phenomenon, and leads to extremely counter-intuitive predictions, which push us to the limit in terms of what is acceptable and compatible with the principle of no action at a distance.

9 Bell's Theorem and Local Hidden Variable Models

9.1 Quick refresher on probabilities and random variables

In this section we will need to use the notion of a random variable, which will require a bit more of probability theory than we have used up until now. In this subsection, we therefore give a quick refresher on the concepts that we will use.

> ***In order to be more concise, we will use*** $P(\lambda)$ ***as a shorthand for*** $\mathrm{Prob}(\lambda)$***, the probability of the event*** λ***.***

Random variables are variables which take on different values probabilistically. The probability that a random variable A takes the value a is $P(a)$.

When we have a pair of random variables, A and B, we will have a ***joint probability distribution*** $P(a,b)$, giving the probability that $A = a$ and $B = b$.

If we are only interested in one of the variables A or B, we use the ***marginal probability distributions*** $P(a)$ or $P(b)$, where

$$P(a) = \sum_b P(a,b), \qquad P(b) = \sum_a P(a,b). \tag{9.1}$$

That is, we sum up all of the probabilities where $A = a$, irrespective of the value of b to obtain the total probability for the event a, and similarly for B.

The ***conditional probability distribution*** arises when the value of one random variables is known, and we want to update the probabilities of the remaining variable, given the new information. We denote by $P(a|b)$ the probability that $A = a$, given that $B = b$. The formula for this update rule is

$$P(a|b) = \frac{P(a,b)}{P(b)}, \tag{9.2}$$

that is, all events are renormalised by dividing by the probability that the event b happened. By combining (9.1) and (9.2) we arrive at the law of total probability

$$P(a) = \sum_b P(a|b)P(b), \tag{9.3}$$

which says that the total probability for $A = a$, is the average of the conditional probabilities that $A = a$ knowing that $B = b$, weighted by the probability that $B = b$.

9.2 Nonlocality of quantum theory

The presence of entangled states makes quantum theory seem like an intrinsically 'nonlocal' theory. In particular, we saw in the previous section, that when measurements are made on half of an entangled state, the other half collapses due to the measurement in a nonlocal way. However, we saw that this doesn't lead to instantaneous observable effects at a distance. Maybe the nonlocality of quantum theory is therefore unnecessary or only apparent.

In fact, maybe this hints at the possibility that there is a ***deeper underlying physical theory*** which is able to reproduce all of the predictions of quantum mechanics but is in fact absolutely ***local***, with no form of nonlocality appearing anywhere in the theory.

To see why this might be plausible, it is instructive to have a think about how random variables behave in classical probability theory, where we can see some analogous effects to entanglement and instantaneous collapse, which challenge us to define exactly what we mean by nonlocality in the first place.

Consider a pair of random variables A and B, which take values $a, b \in \{0, 1\}$ (i.e. a and b are just bits). Let's assume that Alice holds the random variable A and Bob holds B. Let's consider that their associated joint probability distribution is

$$P(a,b) = \begin{cases} \frac{1}{2} & \text{if} \quad a = 0, b = 0, \\ \frac{1}{2} & \text{if} \quad a = 1, b = 1, \\ 0 & \text{otherwise.} \end{cases} \tag{9.4}$$

This corresponds to a situation where Alice and Bob share perfectly correlated bits.

Similarly to quantum theory, there is also an apparent form of nonlocality here: the joint probability distribution cannot be described by giving only the marginal distributions $P(a)$ and $P(b)$, but must be described globally by $P(a, b)$. The marginal distributions are a bit like reduced density operators, and correctly describe the individual probabilities of A and of B, but fail to capture the ***correlations*** between A and B – i.e. the fact that A always takes the same value as B. This looks somewhat akin to entanglement.

Moreover, when Alice learns the value of her random variable, there is a sense in which she 'collapses' to the situation where her random variable takes the observed value with probability 1. Now, if asked what is the probability for Bob's random variable, she will use $P(b|a)$ to describe it – in the present case she will in fact know exactly what Bob will see when he looks at his bit, as it is equal to Alice's bit. Did Bob's bit somehow change when Alice learnt the value of her bit? Obviously not, yet Alice's description indeed changed.

The fact that Alice and Bob have perfectly correlated bits in this example is not at all strange or nonlocal, and can be easily understood. We can think that there was a source which produced A and B, and sent the former to Alice and the latter to Bob. That is, the source produced the pair (a, b) with probability $P(a, b)$, and then sent a to Alice and b to Bob.

The fact that Alice and Bob have perfectly correlated bits can be completely understood in a straightforward way – we say that ***there is a local explanation*** or a ***local model*** which describes this situation. Here, ***locality*** refers to the fact that the correlations were established in the past, in one place, before being distributed to the distant parties.

When the entanglement in quantum mechanics was first identified, many researchers, famously including Einstein, thought that something similar to this classical probabilistic case was all that was really going on. They thought there was an underlying ***local*** theory of physics, and that the nonlocality of quantum theory was completely analogous to the apparent nonlocality of joint probabilities, which is easily understood. This yet-to-be-discovered underlying theory became known as a theory of ***local hidden variables***. The basic elements of this theory were supposed to be some kind of classical random variables which are otherwise 'hidden' from us, but which behave according to the laws of classical probability theory – in particular in a normal local fashion.

However, in 1964 John Bell proved that ***no such local theory can exist*** – the nonlocal weirdness of quantum theory is fundamental and here to stay.

Bell's Theorem:
No theory of local hidden variables can reproduce ***all*** *of the predictions of quantum mechanics.*

9.3 Local hidden variable models

We will only need to consider one specific type of experiment to demonstrate Bell's Theorem. As such, we don't need a fully fledged *theory* of hidden variables, but only a simpler ***model*** which is able to explain just this type of experiment. This is much easier to analyse, and we will refer to this special sub-case as a ***local hidden variable model***.

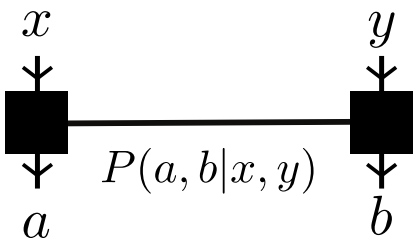

Figure 3: Alice and Bob perform measurements labelled x and y and obtain outcomes a and b. The experiment is characterised by the probabilities $P(a, b|x, y)$.

Consider a situation where Alice and Bob, separated in space, perform a set of experiments, where in each one they locally perform measurements on half an entangled state. Let us abstractly label the measurement made by Alice by x, and that of Bob by y. Similarly, we will always denote the outcomes of Alice's measurements by a, and those of Bob by b.

We need to use this more abstract language as we want to compare the quantum mechanical description of this type of experiment with a local hidden variable model description, and hence we must phrase things in a way which doesn't refer explicitly to the quantum formalism, but only to general, theory-independent aspects.

Example Alice and Bob might share the maximally entangled state

$$|\Phi\rangle = \frac{|0\rangle|0\rangle + |1\rangle|1\rangle}{\sqrt{2}}, \tag{9.5}$$

and Alice might perform in one experiment a measurement of X, and in a second experiment a measurement of Z. We will label the first measurement by $x = 0$, and the second measurement by $x = 1$. In either case, we will call the outcome a, which could be either $+1$ or -1. We will do similar for Bob. In total, we will therefore consider 4 different experiments – one where they both measure X, one where Alice measures X and Bob measures Z, and so forth. In each experiment, there will be four possible ***pairs*** of outcomes – they both obtain $+1$, Alice obtains $+1$ and Bob obtains -1, and so on.

We can collect all of the ***statistics*** produced by such collections of experiments into a set of conditional probabilities

$$P(a, b|x, y)$$

where x and y label the measurements made by Alice, and a and b label the outcomes they receive in each experiment. There are thus 16 probabilities we are interested in, the four possible outcomes of the 4 different experiments.

Returning to our example, the probabilities that will be observed can be calculated using the standard standard rules for performing measurements on quantum states. We shall not perform these calculations here, but simply state the results (which you could easily check). They are

$$\begin{aligned}
P(+1,+1|0,0) = P(-1,-1|0,0) &= \tfrac{1}{2}, \\
P(+1,-1|0,0) = P(-1,+1|0,0) &= 0 \\
P(+1,+1|0,1) = P(+1,-1|0,1) = P(-1,+1|0,1) = P(-1,-1|0,1) &= \tfrac{1}{4} \\
P(+1,+1|1,0) = P(+1,-1|1,0) = P(-1,+1|1,0) = P(-1,-1|1,0) &= \tfrac{1}{4} \\
P(+1,+1|1,1) = P(-1,-1|1,1) &= \tfrac{1}{2}, \\
P(+1,-1|1,1) = P(-1,+1|1,1) &= 0
\end{aligned} \tag{9.6}$$

For example, $P(+1,+1|0,0) = \|(|+\rangle\langle+| \otimes |+\rangle\langle+|)|\Phi\rangle\|^2$ is the probability that both Alice and Bob see the outcome $+1$ when they each measure X on their half of the state.

The probabilities $P(a,b|x,y)$ will be called ***local*** if they can also be reproduced by a local hidden variable model.

Such a model would say that instead of making measurements on quantum systems in order to generate the outcomes a and b, in fact there were random variables sent to Alice and Bob, one corresponding to each measurement. Instead of performing measurements on quantum particles, the model says that there is ***one random variable associated to each measurement***, and ***the value that random variable takes is the result of the measurement***.

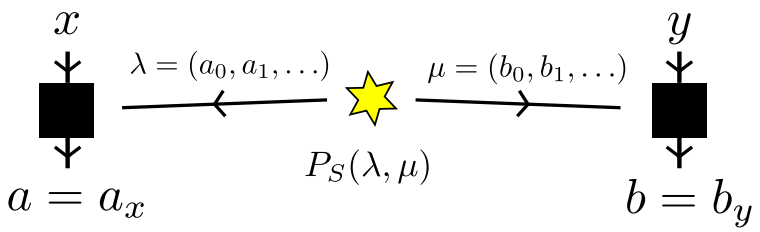

Figure 4: A local hidden variable model. A probabilistic source S emits hidden variables λ to Alice and μ to Bob with probability distribution $P_S(\lambda,\mu)$. $\lambda = (a_0, a_1, \ldots)$ specifies a deterministic list of outcomes, such that $a = a_x$ when Alice thinks she performs the measurement labelled x. Similarly $\mu = (b_0, b_1, \ldots)$ specifies a deterministic list of outcomes, such that $b = b_y$ when Bob thinks he performs the measurement labelled y.

That is, Alice's local hidden variable $\lambda = (a_0, a_1, \ldots)$ is a list of measurement results, one for each choice of x, When the measurement labelled by x is thought to be performed, the result will be given by the value of a_x found on the list, $a = a_x$. Similarly, Bob's hidden variable $\mu = (b_0, b_1, \ldots)$ is a second list of measurement outcomes. In each run of the experiment, Alice may get a different list λ, and Bob may get a different list μ, according to a ***joint*** probability distribution $P_S(\lambda,\mu)$. We say that a pair of lists λ and μ correspond to a ***deterministic strategy***. Crucially, since there is a joint distribution $P_S(\lambda,\mu)$ we can certainly have ***correlations*** between Alice and Bob, but these all have a completely local explanation, arising from this joint distribution of the source.

In the specific situation where Alice receives x and Bob receives y, the (conditional) probabilities that will be generated by such a local hidden variable model will just be the ***average*** over the source,

$$\begin{aligned} P(a,b|x,y) &= \sum_{\lambda,\mu} P_S(\lambda,\mu) P_A(a|x,\lambda) P_B(b|y,\lambda), \\ &= \sum_{\lambda,\mu} P_S(\lambda,\mu) \delta_{a,a_x} \delta_{b,b_y}. \end{aligned} \tag{9.7}$$

where $P_A(a|x,\lambda)$ denotes the probability that Alice's result will be a, when she thought she was making the measurement labelled x, and given that she received λ from the source. Similarly, $P_B(b|y,\mu)$ is the probability that Bob's result will be b, when he thought he was making the measurement labelled by y, and given that he received μ from the source. We obtain the second line, since the model says that if Alice received $\lambda = (a_0, a_1, \ldots)$, then she will definitely give as result $a = a_x$; we have then that

$$P_A(a|x,\lambda) = \begin{cases} 1 & \text{if } a = a_x \\ 0 & \text{if } a \neq a_x \end{cases} = \delta_{a,a_x}, \tag{9.8}$$

and similarly for Bob.

That is, the source S sends the 'results' λ to Alice and μ to Bob, with probability $P_S(\lambda, \mu)$. In Alice's laboratory, instead of making the measurement labelled x, the model gives as result the outcome $a = a_x$ from the list λ, and similarly for Bob.

This form also highlights why we use the terminology ***hidden***: All we observe are a and b. The lists λ and μ are 'hidden' in the background, but contain all the information relevant to the experiments.

We take equation (9.7) as the definition of what it means for the probabilities $P(a, b|x, y)$ to be ***local***. Whenever it is possible to reproduce the statistics generated by a given experiment with a local model, then there is nothing nonlocal about the probabilities $P(a, b|x, y)$.

As an example, the probabilities considered above, when Alice and Bob both measure X and Z on the maximally entangled state are in fact ***local***. The local hidden variable model which is able to perfectly reproduce the probabilities (9.6) is given in Table 1 below:

$P_S(\lambda, \mu)$	Alice λ		Bob μ	
	a_0	a_1	b_0	b_1
1/4	+1	+1	+1	+1
1/4	+1	−1	+1	−1
1/4	−1	+1	−1	+1
1/4	−1	−1	−1	−1

Table 1: Local Hidden Variable Model that is able to reproduce the experimental probabilities in (9.6). An LHV model is specified by a list of deterministic ***stategies*** (λ, μ), with corresponding probabilities $P_S(\lambda, \mu)$. These correspond to the rows of the table. Each strategy specifies each possible measurement outcomes for Alice and Bob. In the above example, the model makes use of 4 deterministic strategies, each with probability 1/4. In the cases when Alice and Bob make use of the first strategy, it says that no matter which measurement each of them think they are making, they will each always return a result of +1. In this way, all of their measurement results are perfectly correlated with each other. However, in other strategies (other rows), we see that they behave differently, and this affects the correlations they observe on average, when using this model.

First, let us look in a little detail at this model to check that it indeed works properly. Each row corresponds to a different strategy, and in this case we see that there are 4 strategies used, each with equal probability $P_S(\lambda, \mu) = 1/4$. This means that the observed behaviour (the statistics $P(a, b|x, y)$) are in fact an average of 4 distinct (deterministic) behaviours.

In the first strategy (the first row), the result of the measurement of X is +1 (a_0), and the result of the measurement of Z is also +1 (a_1). Similarly, for both of Bob's measurements, both results are +1 (b_0 and b_1). If Alice and Bob would use ***only*** this strategy, the associated probabilities they would observe would be

$$P(+1, +1|0, 0) = P(+1, +1|0, 1) = P(+1, +1|1, 0) = P(+1, +1|1, 1) = 1, \tag{9.9}$$

and $P(a, b|x, y) = 0$ otherwise. This is obviously different from (9.6), and shows that using only the first strategy all of the time would definitely not reproduce the measurement results on the maximally entangled state.

For the 3 other rows, corresponding to 3 different strategies, if we also consider them individually, they would lead to the following probabilities: For the strategy of row 2:

$$P(+1, +1|0, 0) = P(+1, -1|0, 1) = P(-1, +1|1, 0) = P(-1, -1|1, 1) = 1; \tag{9.10}$$

For the strategy of row 3:

$$P(-1, -1|0, 0) = P(-1, +1|0, 1) = P(+1, -1|1, 0) = P(+1, +1|1, 1) = 1; \tag{9.11}$$

For the strategy of row 4:

$$P(-1,-1|0,0) = P(-1,-1|0,1) = P(-1,-1|1,0) = P(-1,-1|1,1) = 1; \tag{9.12}$$

In all cases, all other probabilities $P(a,b|x,y)$ vanish, and in each case we clearly see that we don't produce the measurement results we desire in (9.6).

To calculate the statistics that the model as a whole leads to, we therefore need to average over these 4 different strategies, each of which is used with probability $P_S(\lambda,\mu)$. In our present example, each strategy is used $1/4$ of the time. That is, here we can simply sum up (9.9) – (9.12), and divide by 4. If we carry out this calculation, we see that it precisely reproduces the probabilities in (9.6). This demonstrates that this model is indeed able to reproduce these probabilities.

How did we arrive at this model? There are two steps that we can take to get there. The first step is to rule out certain strategies, that is, to show that there are certain strategies that we definitely cannot use. We can do this by looking at those probabilities that are ***zero*** in $P(a,b|x,y)$. In our example (9.6), there were 4 probabilities that are zero:

$$P(+1,-1|0,0) = P(-1,+1|0,0) = P(+1,-1|1,1) = P(-1,+1|1,1) = 0.$$

These zero probabilities mean that some strategies cannot be used. For example, focusing on the first of these, $P(+1,-1|0,0) = 0$, imagine any strategy where $a_0 = +1$ and $b_0 = -1$. If we were to use such a strategy with certainty, it would predict $P(+1,-1|0,0) = 1$, which is not what we want to reproduce. If we were to use it with some non-zero probability P_S, then we would still have $P(+1,-1|0,0) > 0$, which is inconsistent with what we want to reproduce. This shows that we can never use any strategy where $a_0 = +1$ ***and*** $b_0 = -1$, and all strategies where this is the case can be discounted. This rules out the possibility of using 4 strategies of the form $\lambda = (a_0,+1)$, $\mu = (b_0,-1)$, for any value of a_0 or b_0.

Similarly, focusing on the second, third and fourth zeros, any strategies with $a_0 = -1$ and $b_0 = +1$ can be discounted, as can strategies with $a_1 = +1$ and $b_1 = -1$, and strategies with $a_1 = -1$ and $b_1 = +1$. A little thought, and experimentation, shows that the ***only*** strategies that are not discounted are the four appearing in Table 1. That is, these 4 strategies are the only 4 that are not discounted just from the zeros in $P(a,b|x,y)$ in (9.6).

Second, we can also figure out how ***many*** strategies are necessary to use, and their corresponding probabilities. In our example, we see that for $x = 0$, $y = 1$, all 4 probabilities occur uniformly, $P(a,b|0,1) = 1/4$ for all a, b. Now, any individual strategy leads to probabilities of the form (9.9), where only one outcome occurs, with probability 1. In order to have 4 non-zero probabilities, we therefore need to make use of ***at least*** 4 strategies, covering all the different pair of values that a_0 and b_1 can jointly take. If we look at our model, this is exactly what we have. Moreover, we actually only had 4 strategies left to make use of anyway, so we couldn't have made use of more than 4, but now we also see that we cannot take less than 4 either. This shows that we have to make use of all 4 strategies, and that the probabilities of each have to be $1/4$ in order to correctly reproduce $P(a,b|0,1)$.

What the above shows is that the probabilities (9.6), although they arose from measurements on a ***maximally entangled state***, can also be explained using a local hidden variable model. As such, they don't demonstrate any form of nonlocality in quantum mechanics.

If this would have been true for ***any experiment*** of this form, then we may have started to believe that a ***theory*** of local hidden variables exists that could potentially reproduce all of the predictions of quantum mechanics. At the very least, we would have concluded that nothing strange or unusual was going on in this very special types of experiment, which involve entanglement, and local measurements collapsing the state.

However, as we will show in the next section, ***not all probabilities that arise from measurements on entangled quantum states in such experiments can be reproduced by a local hidden variable model***. That is, we are going to show – remarkably – that we can find probabilities $P(a,b|x,y)$

arising from measurements on entangled quantum states, such that if we try and find a local hidden variable model that can reproduce them, that this is simply ***impossible***.

This is precisely Bell's theorem, now stated in a more exact way.

The challenge however is – ***how can we possibly convince ourselves that no local hidden variable model exists that can reproduce a given set of probabilities?*** This seems like a daunting challenge. Luckily, there is an ingenious way to acheive this! We will carefully identify ***limitations*** that local hidden variable models must obey. If we can then show that quantum mechanics ***doesn't obey*** these limitations, then it means that we ***definitely can't find a local hidden variable model*** reproducing the quantum behaviour. These limitations are also wonderfully intuitive, and correspond to ***limitations on how well we can play certain cooperative 'games'***, as we will see in the next section.

9.4 The CHSH game

The proof of Bell's theorem given here will be based around a simpler and more powerful version, obtained shortly after Bell's original proof, by Clauser, Horne, Shimony and Holt. The presentation given is from a modern quantum information perspective, which presents it in terms of a 'game'.

We are going to show that Alice's and Bob's probability of success of winning a carefully constructed co-operative game is limited below a certain threshold,whenever we assume that their physics is described by a local hidden variable theory. On the other hand, in the next section, we will show that by using quantum theory – i.e. by utilising a strategy to play the game that consists of performing measurements on an entangled quantum state – in order to help play the game, a strictly larger success probability can be achieved!

This will prove conclusively that the predictions of quantum theory are inconsistent with any underlying physical theory which is described by local hidden variables.

The CHSH game (non-mathematical version):

1. One weekend, Alice and Bob each have 50% chance of being visited by a friend (Amy and Bill respectively) on Saturday.
2. Alice and Bob would like to meet up on Saturday afternoon, but must avoid Amy and Bill from meeting (since they are not on speaking terms).
3. Alice and Bob meet on Thursday to decide in advance what to do (to devise a strategy). They can't communicate during the weekend (to avoid suspicion from Amy or Bill).
4. Alice and Bob must each decide on Saturday afternoon – after Amy and Bill have arrived – between either going to the pub or to the park on Saturday afternoon.
5. They 'win' if they meet up when it is safe to do so (one of fewer friends visit) or avoid each other when both friends visit; if they avoid each other when it was safe to meet, or meet up when they shouldn't have, they lose the game.

What is the maximum probability p^{suc} of achieving a successful outcome? That is, of Alice and Bob meeting up when it is safe to do so, and avoiding each other when both friends happen to visit at the same time? The difficulty is that on Saturday morning, when she needs to decide where to go, Alice will not know if Bill has come to visit Bob or not. Similarly, Bob will not know if Amy has come to visit Alice or not when he needs to decide where to go with Bill on Saturday afternoon.

One plan that Alice and Bob could agree upon on Thursday is that Alice will always go to the

pub, irrespective of whether Amy visits or not, and Bob goes to the park if Bill visits, and to the pub if he doesn't visit. This plan only works in 3 of the 4 possible scenarios, since if Amy doesn't visit, Bob and Bill should meet up with Alice, and should therefore go to the pub, yet they go to the park and don't meet up with Alice. Since all four scenarios are equally likely, this means we can think that this plan achieves $p^{\text{suc}} = 3/4$, as it only works in 3 out of 4 scenarios (we could alternatively think of repeating this many times, or having many Alice's and Bobs, and looking at the fraction of times when they do the right thing versus doing the wrong thing).

If Alice and Bob only have access to 'local' physics (i.e. to a local hidden variable model), there is no better plan of action than this:

$$p^{\text{suc}}_{\text{local}} \leq \frac{3}{4}. \tag{9.13}$$

To prove this, we will use a ***mathematical reformulation*** of the CHSH game:

The CHSH game (mathematical version):

1. Alice and Bob receive bits $x \in \{0,1\}$, $y \in \{0,1\}$ at random: $P(x,y) = 1/4$.
2. Alice and Bob must produce bits $a \in \{0,1\}$ and $b \in \{0,1\}$ respectively.
3. Alice and Bob win if the bits they produce satisfy the relation

$$a \oplus b = xy \tag{9.14}$$

   where $\oplus$ is ***addition modulo 2*** (see below)

4. The success probability is given by

$$\begin{aligned} P^{\text{suc}} &= \sum_{x,y} P(x,y) P(a \oplus b = xy | x, y) \\ &= P(a \oplus b = xy) \end{aligned} \tag{9.15}$$

   where $P(a \oplus b = xy|x,y)$ is shorthand for

$$\begin{aligned} &P(0,0|x,y) + P(1,1|x,y) \quad \text{if } xy = 0 \\ &P(0,1|x,y) + P(1,0|x,y) \quad \text{if } xy = 1. \end{aligned}$$

Addition modulo 2 satisfies

$$\begin{aligned} 0 \oplus 0 &= 0 \\ 0 \oplus 1 &= 1 \\ 1 \oplus 0 &= 1 \\ 1 \oplus 1 &= 0 \end{aligned}$$

and can be thought of as the natural way to define addition when there are only 2 numbers, 0 and 1 (if this seems strange, just think about a clock, which is addition modulo 12. 3 hours after 10am is 1pm – we are so used to performing addition modulo 12 we don't even think about it. Just as this is most easily represented by placing the numbers on the circle of the clock face, addition modulo 2 can be easily visualised as the 2 numbers 0 and 1 in a circle).

To understand this mathematical formulation, we start by identifying the elements: x represents whether Amy comes to visit or not: $x = 0$ means she doesn't come to visit, while $x = 1$ means she does. Bill is similarly modelled by the bit y. a then represents where Alice goes: $a = 0$ represents going to the pub, and $a = 1$ represents the park; the same is true for b. What remains is to understand why $a \oplus b = xy$ represents them doing the correct thing. On the one hand, $a \oplus b = 0$

if $a = b$, so this is a way of checking that they went to the same place; similarly $a \oplus b = 1$ means $a \neq b$, and hence they means they went to different places. The product $xy = 0$ if either $x = 0$ or $y = 0$, i.e. if either of the friends doesn't visit (or neiter visits) then $xy = 0$, and so $a \oplus b = 0$, and hence Alice and Bob should go to the same place; on the other hand if both friends visit, $xy = 1$, and so the $a \oplus b = 1$ tells us they should go to different places to win.

If Alice and Bob can only use local hidden variables to play the CHSH game, then they can never achieve $p^{\text{suc}} = 1$. The most general local hidden variable model they can use has the form

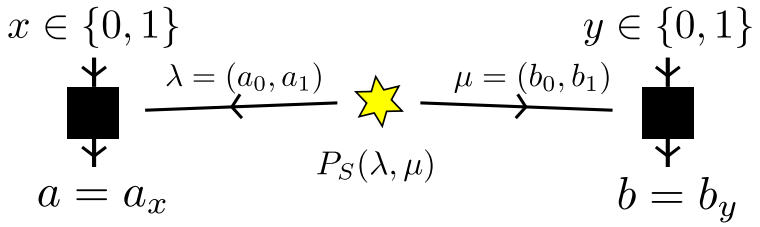

Figure 5: General local hidden variable strategy for playing the CHSH game.

where $\lambda = (a_0, a_1)$ and $\mu = (b_0, b_1)$ are lists of responses for Alice and Bob, where Alice will give as an outcome $a = a_0$, when $x = 0$ and so on. This is just the mathematical way of referring to the plan that Alice and Bob make on Thursday night, for example, it is the mathematical way of saying that Alice decides she will go to the pub if Amy comes to visit, but will go to the park if she doesn't.

Let us assume first for simplicity that only a single λ and μ are used by Alice and Bob. That is, we assume that the source of hidden variables is ***deterministic*** – the probability distribution $P_S(\lambda, \mu)$ is equal to 1 for some specific choice of $\lambda = (a_0, a_1)$ and $\mu = (b_0, b_1)$, and 0 otherwise. This would correspond to Alice and Bob always deciding on the same fixed plan of action in the above informal form with Amy and Bill. If the source was probabilistic, this would correspond to them deciding to act probabilistically, for example by flipping coins on Thursday in order to decide what to do.

For such a ***deterministic local hidden variable model*** to achieve $p^{\text{suc}} = 1$, then the following four conditions must necessarily hold

$$\begin{aligned} a_0 \oplus b_0 &= 0, \\ a_0 \oplus b_1 &= 0, \\ a_1 \oplus b_0 &= 0, \\ a_1 \oplus b_1 &= 1. \end{aligned} \tag{9.16}$$

which is just saying that in the 4 possible scenarios, where $x = 0$ and $y = 0$, $x = 0$ and $y = 1$, and so on, the answers of Alice and Bob, which will be $a = a_x$ and $b = b_x$, must be correct.

Now, we want to show that there is no possible assignment of values for a_0, a_1, b_0 and b_1 that simultaneously satisfy these four equations. Although they may look different to what we are familiar with, this is still four simultaneous equations in four unknowns, and so we should be able to either find a solution, or show that no solution can occur. Let's proceed by 'adding' the equations together. In this context, we should be careful to add them up ***modulo 2***, in which case we find

$$\begin{aligned} & (a_0 \oplus b_0) \oplus (a_0 \oplus b_1) \oplus (a_1 \oplus b_0) \oplus (a_1 \oplus b_1) = 0 \oplus 0 \oplus 0 \oplus 1, \\ \implies \; & (a_0 \oplus a_0) \oplus (a_1 \oplus a_1) \oplus (b_0 \oplus b_0) \oplus (b_1 \oplus b_1) = 1, \\ & \implies 0 = 1. \end{aligned} \tag{9.17}$$

To obtain the second line, we simply simplified the right-hand side, and re-bracketed on the left-hand side, and to obtain the third we used the fact that $c \oplus c = 0$ for any bit c (because c is always

equal to c).

The fact that we arrived at something absurd – that $0 = 1$, implies that it is impossible that all four equations can be simultaneously satisfied.

This shows that any ***deterministic*** model (with some fixed λ and fixed μ) must lose the game for at least one pair of values for x and y. The probability of success for any λ, μ is therefore bounded

$$P^{\text{suc}}(\lambda, \mu) = P(a \oplus b = xy|\lambda, \mu) \leq \frac{3}{4}. \tag{9.18}$$

More generally, we can allow the source to distribute different λ and μ with probability $P_S(\lambda, \mu)$. However, it is easy to verify that this in fact cannot help. In particular,

$$\begin{aligned} P^{\text{suc}}_{\text{local}} &= \sum_{\lambda,\mu} P_S(\lambda, \mu) P^{\text{suc}}(\lambda, \mu), \\ &\leq \sum_{\lambda,\mu} P_S(\lambda, \mu) \frac{3}{4}, \\ &= \frac{3}{4}. \end{aligned} \tag{9.19}$$

where we have denoted by $P^{\text{suc}}_{\text{local}}$ the success probability assuming a ***local hidden variable model***. This statement is very intuitive – if, no matter what plan of action is chosen, the game cannot be won more than $3/4$ of the time, it won't suddenly become possible to win more frequently just by mixing such plans together.

Altogether, this proves that ***the success probability of winning the CHSH game is upper bounded by*** $3/4$ ***assuming that physics is described by local hidden variables.*** This thus constitutes a ***non-trivial*** limitations that is satisfied by all local hidden variable models.

In the next section we will see that it is possible to win the CHSH game with a much larger probability (approximate 85% of the time) by using a quantum strategy involving measurements on shared entangled states to play the game. This then demonstrates conclusively that quantum mechanics can produce correlations in this simple situation that cannot be reproduced (or explained) by a local hidden variable model.

10 Quantum nonlocality

In this section, we will see that, remarkably, if Alice and Bob employ a ***quantum strategy*** to play the CHSH game, that this can actually help them win with a higher success probability than is possible using only local hidden variables. Thinking to the informal version of the CHSH game, this means that if Alice and Bob choose where to go by making measurements on quantum systems – an idea which are first sight sounds rather absurd – this will actually help them co-ordinate better than they could possibly do by just planning in advance! Therefore, even though there is no signalling in quantum mechanics, making measurements on entangled particles allows for levels of co-ordination which are simply inexplicable from a classical perspective.

The quantum strategy that they will employ is to produce an entangled pair of qubits, of which Alice keeps one, and Bob keeps the other. When they get given the inputs to the CHSH game (x and y respectively), they use them to choose between two measurements which they perform on their half of the state. They then use the outcomes of the measurement to produce the answers to the game, a and b. We will see in this section that by judiciously choosing a state and measurements, that such a quantum strategy for the CHSH game allows them to win with probability

$$P^{\text{suc}}_{\text{quantum}} = \frac{2+\sqrt{2}}{4} \approx 0.85, \tag{10.1}$$

which is substantially higher than the $3/4$ probability with which they could have won using only a local hidden variable model as the basis for their strategy.

10.1 The quantum strategy

A more detailed overview of the quantum strategy that we will consider is the following:

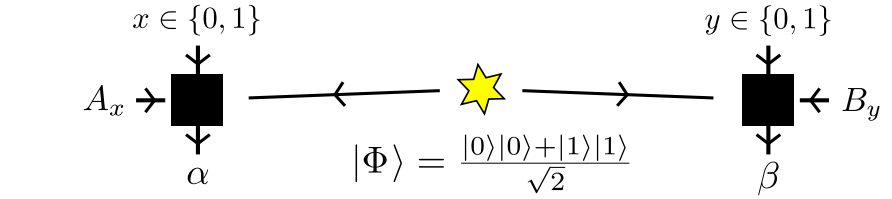

Figure 6: Quantum strategy that Alice and Bob can use to play the CHSH game. They shared the maximally entangled state $|\Phi\rangle$, and when given input x and y, make measurements A_x and B_y, which produce outcomes $\alpha, \beta \in \{+1, -1\}$. They use these outcomes to produce the answers to the game a and b.

That is, Alice and Bob share the maximally entangled state of two qubits,

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

Alice measures one of the operators A_0 or A_1, depending on whether she receives as input $x = 0$ or $x = 1$. Both A_0 and A_1 have eigenvalues $\alpha \in \{+1, -1\}$. Similarly, Bob measures either B_0 or B_1, depending on whether he receives $y = 0$ or $y = 1$, which again both have eigenvalues $\beta \in \{+1, -1\}$. In order to produce the outcomes a and b, Alice and Bob use the mapping

$$a = \begin{cases} 0 & \text{if} \quad \alpha = +1, \\ 1 & \text{if} \quad \alpha = -1. \end{cases} \qquad b = \begin{cases} 0 & \text{if} \quad \beta = +1, \\ 1 & \text{if} \quad \beta = -1. \end{cases} \tag{10.2}$$

The operators that Alice and Bob measure are combinations of X and Z. The measurements can be conveniently depicted as in Figure 7 (which corresponds to the equator of the Bloch sphere):

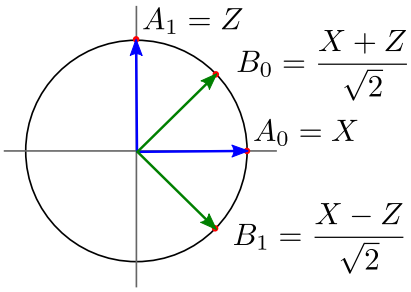

Figure 7: Operators that Alice and Bob should measure to play the CHSH game. The direction of the arrow representing each operator corresponds to the angle θ in $S(\theta)$, where we go clockwise starting from the y axis.

To better understand these operators, consider a more general class of operators, of which they are special cases:

$$S(\theta) = \cos(\theta)Z + \sin(\theta)X. \tag{10.3}$$

Such operators are Hermitian for all θ, and can be shown to have eigenvalues $+1$ and -1 independent of θ. The corresponding eigenstates are in fact

$$|\uparrow_\theta\rangle = \cos\tfrac{\theta}{2}|0\rangle + \sin\tfrac{\theta}{2}|1\rangle, \tag{10.4}$$
$$|\downarrow_\theta\rangle = -\sin\tfrac{\theta}{2}|0\rangle + \cos\tfrac{\theta}{2}|1\rangle, \tag{10.5}$$

as can be straightforwardly checked by direct substitution. As such, the spectral decomposition of $S(\theta)$ is given by

$$S(\theta) = |\uparrow_\theta\rangle\langle\uparrow_\theta| - |\downarrow_\theta\rangle\langle\downarrow_\theta| \tag{10.6}$$

which we now verify,

$$\begin{aligned} S(\theta) &= |\uparrow_\theta\rangle\langle\uparrow_\theta| - |\downarrow_\theta\rangle\langle\downarrow_\theta|, \\ &= (\cos\tfrac{\theta}{2}|0\rangle + \sin\tfrac{\theta}{2}|1\rangle)(\cos\tfrac{\theta}{2}\langle 0| + \sin\tfrac{\theta}{2}\langle 1|) - (-\sin\tfrac{\theta}{2}|0\rangle + \cos\tfrac{\theta}{2}|1\rangle)(-\sin\tfrac{\theta}{2}\langle 0| + \cos\tfrac{\theta}{2}\langle 1|), \\ &= (\cos^2\tfrac{\theta}{2} - \sin^2\tfrac{\theta}{2})(|0\rangle\langle 0| - |1\rangle\langle 1|) + 2\sin\tfrac{\theta}{2}\cos\tfrac{\theta}{2}(|0\rangle\langle 1| + |1\rangle\langle 0|), \\ &= \cos(\theta)Z + \sin(\theta)X. \end{aligned} \tag{10.7}$$

We see that the operators that we introduced above in Fig. 7 are all from this family of operators:

$$\begin{aligned} A_0 &= X = S\left(\tfrac{\pi}{2}\right), & B_0 &= \frac{X+Z}{\sqrt{2}} = S\left(\tfrac{\pi}{4}\right), \\ A_1 &= Z = S\left(0\right), & B_1 &= \frac{X-Z}{\sqrt{2}} = S\left(\tfrac{3\pi}{4}\right), \end{aligned} \tag{10.8}$$

hence all of Alice's and Bob's outcomes $\alpha, \beta \in \{+1, -1\}$, as claimed.

It remains now to calculate the success probability of winning the CHSH game using this quantum strategy. Written out in full, the success probability is

$$\begin{aligned} P^{\text{suc}} &= \frac{1}{4}\sum_{x,y} P(a \oplus b = xy|x,y), \\ &= \frac{1}{4}\Big(P(a\oplus b = 0|x=0,y=0) + P(a\oplus b = 0|x=0,y=1) \\ &\qquad + P(a\oplus b = 0|x=1,y=0) + P(a\oplus b = 1|x=1,y=1)\Big). \end{aligned} \tag{10.9}$$

What we will first show is that the probabilities appearing in this expression, when using the above quantum strategy, are given by,

$$P(a \oplus b = 0|x,y) = \tfrac{1}{2}(1 + \langle\Phi|A_x \otimes B_y|\Phi\rangle), \tag{10.10}$$
$$P(a \oplus b = 1|x,y) = \tfrac{1}{2}(1 - \langle\Phi|A_x \otimes B_y|\Phi\rangle), \tag{10.11}$$

for all $x = 0$ or 1 and $y = 0$ or 1. These expressions are helpful, as they reduce our problem to one of simply calculating the expectation values $\langle\Phi|A_x \otimes B_y|\Phi\rangle$.

To prove the above relation, we must recall that quantum mechanical expectation values coincide with classical expectation values of the measurement results. In particular, for *any* state $|\Psi\rangle$ and *any* pair of operators A and B with eigenvalues $\alpha = \pm 1$ and $\beta = \pm 1$ respectively, it holds ***by definition*** that

$$\langle\Psi|A \otimes B|\Psi\rangle = \sum_{\alpha,\beta} \alpha\beta \,\mathrm{Prob}(\alpha,\beta) \tag{10.12}$$

In our case, we further see that

$$\begin{aligned}
\langle\Psi|A \otimes B|\Psi\rangle &= \sum_{\alpha,\beta} \alpha\beta \,\mathrm{Prob}(\alpha,\beta), \\
&= \mathrm{Prob}(\alpha = +1, \beta = +1) - \mathrm{Prob}(\alpha = +1, \beta = -1) \\
&\quad - \mathrm{Prob}(\alpha = -1, \beta = +1) + \mathrm{Prob}(\alpha = -1, \beta = -1), \\
&= \mathrm{Prob}(\alpha = \beta) - \mathrm{Prob}(\alpha \neq \beta), \\
&= \mathrm{Prob}(\alpha = \beta) - (1 - \mathrm{Prob}(\alpha = \beta)), \\
&= 2\,\mathrm{Prob}(\alpha = \beta) - 1.
\end{aligned} \tag{10.13}$$

In the third line we have defined

$$\mathrm{Prob}(\alpha = \beta) = \mathrm{Prob}(\alpha = +1, \beta = +1) + \mathrm{Prob}(\alpha = -1, \beta = -1)$$

as the total probability that α and β are the same, and similarly

$$\mathrm{Prob}(\alpha \neq \beta) = \mathrm{Prob}(\alpha = +1, \beta = -1) + \mathrm{Prob}(\alpha = -1, \beta = +1)$$

as the total probability that α and β are different, which, by definition, add up to one.

Note that to actually calculate $\mathrm{Prob}(\alpha,\beta)$, we need to use the standard quantum mechanical rule – by using the eigenstates of A and B corresponding to the eigenvalues α and β, and by looking at the norm of the state after projecting onto these eigenstates. In the above, we did not need to actually calculate these probabilities, and simply wrote them abstractly as $\mathrm{Prob}(\alpha,\beta)$.

By re-arranging (10.12), we arrive at

$$\mathrm{Prob}(\alpha = \beta) = \tfrac{1}{2}\left(1 + \langle\Psi|A \otimes B|\Psi\rangle\right). \tag{10.14}$$

Finally, to connect to Alice's and Bob's quantum strategy for playing CHSH, we recall that Alice's strategy is to give as output $a = 0$ when the result of her measurement is $\alpha = +1$, and $a = 1$ when the result is $\alpha = -1$, and similarly for Bob in order to turn his result β into his outcome b. Thus, $\alpha = \beta$ implies directly that $a = b$, since the latter will be equal only when the former are. Now, $a = b$ whenever $a = 0$ and $b = 0$ or $a = 1$ and $b = 1$. In both of these cases (and *only* in these cases) we see that $a \oplus b = 0$. That is, a completely equivalent way of writing $a = b$ when a and b are bits is to write $a \oplus b = 0$.

Putting everything together, in the situation where Alice receives x and Bob receives y, Alice measures the operator A_x and Bob measures B_y, whenever their outcomes α and β are equal, then the bits they produce will satisfy $a \oplus b = 0$. Equation (10.14) gave the total probability that the

outcomes of two measurements will be equal for a generic state $|\Psi\rangle$ and generic operators A and B. For the specific case we are interested in, it implies that

$$P(a \oplus b = 0|x, y) = \tfrac{1}{2}(1 + \langle\Phi|A_x \otimes B_y|\Phi\rangle) \tag{10.15}$$

We can perform a similar analysis for the case where $\alpha \neq \beta$. Using the fact that $\text{Prob}(\alpha = \beta) + \text{Prob}(\alpha \neq \beta) = 1$ and (10.14) we arrive at

$$\text{Prob}(\alpha \neq \beta) = \tfrac{1}{2}\left(1 - \langle\Psi|A \otimes B|\Psi\rangle\right). \tag{10.16}$$

As above, given Alice's and Bob's strategy, if $\alpha \neq \beta$ then $a \neq b$. This means either that $a = 0$ and $b = 1$ or $a = 1$ and $b = 0$, and in both cases $a \oplus b = 1$. Hence, a completely equivalent way of writing $a \neq b$ for bits is to write $a \oplus b = 1$. Putting everything together, in the same way as above, we thus arrive at

$$P(a \oplus b = 1|x, y) = \tfrac{1}{2}(1 - \langle\Phi|A_x \otimes B_y|\Phi\rangle) \tag{10.17}$$

which proves the second of the claimed relations we wanted to prove.

What remains is to calculate the expectation values of the form $\langle\Phi|A_x \otimes B_y|\Phi\rangle$. Recalling that all of the operators A_0, A_1, B_0 and B_1 are from the family $S(\theta)$, if we find a general expression for $\langle\Phi|S(\theta) \otimes S(\phi)|\Phi\rangle$ then we will be able to find the expectation value in the 4 cases we are ultimately interested in by substituting the appropriate values for θ and ϕ. What we will now show is that

$$\langle\Phi|S(\theta) \otimes S(\phi)|\Phi\rangle = \cos(\theta - \phi). \tag{10.18}$$

Using the definition of $S(\theta)$, we see that

$$\begin{aligned}\langle\Phi|S(\theta) \otimes S(\phi)|\Phi\rangle &= \langle\Phi|(\cos(\theta)Z + \sin(\theta)X) \otimes (\cos(\phi)Z + \sin(\phi)X)|\Phi\rangle, \\ &= \cos\theta\cos\phi\langle\Phi|Z \otimes Z|\Phi\rangle + \cos\theta\sin\phi\langle\Phi|Z \otimes X|\Phi\rangle \\ &\quad + \sin\theta\cos\phi\langle\Phi|X \otimes Z|\Phi\rangle + \sin\theta\sin\phi\langle\Phi|X \otimes X|\Phi\rangle. \end{aligned} \tag{10.19}$$

We must thus evaluate 4 expectation values, involving only X and Z. In particular, we can easily check that $(X \otimes X)|\Phi\rangle = |\Phi\rangle$ and $(Z \otimes Z)|\Phi\rangle$ and thus

$$\langle\Phi|X \otimes X|\Phi\rangle = \langle\Phi|\Phi\rangle = 1, \tag{10.20}$$
$$\langle\Phi|Z \otimes Z|\Phi\rangle = \langle\Phi|\Phi\rangle = 1. \tag{10.21}$$

Similarly, $(X \otimes Z)|\Phi\rangle = \frac{|1\rangle|0\rangle - |0\rangle|1\rangle}{\sqrt{2}} = |\Psi'\rangle$ and $(Z \otimes X)|\Phi\rangle = \frac{|0\rangle|1\rangle - |1\rangle|0\rangle}{\sqrt{2}} = -|\Psi'\rangle$, and thus

$$\langle\Phi|X \otimes Z|\Phi\rangle = \langle\Phi|\Psi'\rangle = 0, \tag{10.22}$$
$$\langle\Phi|Z \otimes X|\Phi\rangle = -\langle\Phi|\Psi'\rangle = 0. \tag{10.23}$$

where we used the fact that $|\Psi'\rangle$ is orthogonal to $|\Phi\rangle$ and hence $\langle\Phi|\Psi'\rangle = 0$. Substituting these results into (10.19), we arrive at

$$\begin{aligned}\langle\Phi|S(\theta) \otimes S(\phi)|\Phi\rangle &= \sin\theta\sin\phi + \cos\theta\cos\phi, \\ &= \cos(\theta - \phi). \end{aligned} \tag{10.24}$$

Before moving on, let us think a little about this formula. If Alice and Bob measure the same operator, then $\theta = \phi$ and $\cos(\theta - \phi) = \cos(0) = 1$. From (10.14), this then means that $\alpha = \beta$ with certainty, i.e. Alice and Bob always obtain the ***same*** outcome. In (9.6) we already observed that this was the case when Alice and Bob both measure Z or both measure X on $|\Phi\rangle$. What the above shows, is that this is actually true for any operator in the family $S(\theta)$. It also tells us that if Alice and Bob don't measure exactly the same operators, but ones which are "similar", i.e. such that $\theta \approx \phi$, then $\cos(\theta - \phi) \approx 1$, and hence $P(\alpha = \beta) \approx 1$, meaning that they still obtain the same

outcomes most of the time. This is a very special property of $|\Phi\rangle$ and is the key to Alice's and Bob's quantum strategy.

We now have everything we need in order to analyse the success probability of Alice's and Bob's quantum strategy. Let us proceed case by case.

If Alice receives $x = 0$ and Bob receives $y = 0$, then Alice measures $A_0 = S(\frac{\pi}{2}) = X$ and Bob measures $B_0 = S(\pi/4) = \frac{X+Z}{\sqrt{2}}$. The probability of them correctly producing bits a and b that satisfy $a \oplus b = xy = 0$, from (10.15) is

$$\begin{aligned} P(a \oplus b = 0|x = 0, y = 0) &= \tfrac{1}{2}(1 + \langle\Phi|S\left(\tfrac{\pi}{2}\right) \otimes S\left(\tfrac{\pi}{4}\right)|\Phi\rangle), \\ &= \tfrac{1}{2}(1 + \cos\left(\tfrac{\pi}{2} - \tfrac{\pi}{4}\right)), \\ &= \tfrac{1}{2}(1 + \tfrac{1}{\sqrt{2}}), \\ &\approx 0.854, \end{aligned} \tag{10.25}$$

where we used the fact that $\cos(\frac{\pi}{4}) = \frac{1}{\sqrt{2}}$.

Thus they win in this instance approximately 85% of the time. Thinking back to the informal version with Amy and Bill, this means that if neither friend came to visit, than Alice and Bob will both end up in the same place (either the park of the pub), approximately 85% of the time. Note that we didn't calculate here whether they go to the pub or to the park – we of course could also calculate this, but right now we avoided making such calculations, and directly calculated just the total probability that they went to the same place.

In the case where Alice receives $x = 0$ and Bob receives $y = 1$, then the only thing that changes relative to the above is that Bob measures instead $B_1 = S(\frac{3\pi}{4}) = \frac{X-Z}{\sqrt{2}}$. The probability that they produce bits satisfying $a \oplus b = xy = 0$ is now

$$\begin{aligned} P(a \oplus b = 0|x = 0, y = 1) &= \tfrac{1}{2}(1 + \langle\Phi|S\left(\tfrac{\pi}{2}\right) \otimes S\left(\tfrac{3\pi}{4}\right)|\Phi\rangle), \\ &= \tfrac{1}{2}(1 + \cos\left(\tfrac{\pi}{2} - \tfrac{3\pi}{4}\right)), \\ &= \tfrac{1}{2}(1 + \tfrac{1}{\sqrt{2}}), \\ &\approx 0.854, \end{aligned} \tag{10.26}$$

where we used the fact that $\cos(-\frac{\pi}{4}) = \frac{1}{\sqrt{2}}$. Thus Alice and Bob produce equal bits with exactly the same probability as in the previous case.

In the case $x = 1$ and $y = 0$, now Alice measures $A_1 = S(0) = Z$ and Bob measures $B_0 = S(\frac{\pi}{4}) = \frac{X+Z}{\sqrt{2}}$, and we are interested in the probability that $a \oplus b = xy = 0$,

$$\begin{aligned} P(a \oplus b = 0|x = 1, y = 0) &= \tfrac{1}{2}(1 + \langle\Phi|S\left(0\right) \otimes S\left(\tfrac{\pi}{4}\right)|\Phi\rangle), \\ &= \tfrac{1}{2}(1 + \cos\left(-\tfrac{\pi}{4}\right)), \\ &= \tfrac{1}{2}(1 + \tfrac{1}{\sqrt{2}}) \\ &\approx 0.854 \end{aligned} \tag{10.27}$$

Up until this stage, we see that Alice and Bob have found a strategy which (in the informal setting) has them going to the same place with high probability, around 85% of the time. If we think back to our best strategy using local hidden variables, we might now expect that our strategy for Alice and Bob will perform badly in the final case. Indeed, using local hidden variables, whenever Alice and Bob did the correct thing in 3 cases, they ***always*** did the wrong thing in the final case. Although now Alice and Bob behave probabilistically (using their measurement result to decide where to go), we might nevertheless expect that they are likely to do the wrong thing in the final case.

When $x = 1$ and $y = 1$, we would like Alice and Bob to go to opposite places, which was equivalent in the mathematical version of CHSH that they produce bits that satisfy $a \oplus b = xy = 1$.

Using their quantum strategy, from (10.17), we see that the probability of them outputting the correct bits is

$$\begin{aligned} P(a \oplus b = 1 | x = 1, y = 1) &= \tfrac{1}{2}(1 - \langle \Phi | S(0) \otimes S\left(\tfrac{3\pi}{4}\right) | \Phi \rangle), \\ &= \tfrac{1}{2}(1 - \cos\left(-\tfrac{3\pi}{4}\right)), \\ &= \tfrac{1}{2}(1 + \tfrac{1}{\sqrt{2}}) \\ &\approx 0.854 \end{aligned} \tag{10.28}$$

where we used the fact that $\cos(-\frac{3\pi}{4}) = -\frac{1}{\sqrt{2}}$. We notice that this minus sign cancels the minus sign which arose from considering $a \oplus b = 1$ instead of $a \oplus b = 0$, and hence once again Alice and Bob do the correct thing approximately 85% of the time!

So, in all 4 cases, if Alice and Bob employ the quantum strategy, they do the correct thing approximately 85% of the time. Their quantum success probability is thus

$$\begin{aligned} P^{\text{suc}}_{\text{quantum}} &= \frac{1}{4} \sum_{x,y} P(a \oplus b = xy | x, y), \\ &= \tfrac{1}{2}(1 + \tfrac{1}{\sqrt{2}}) = \tfrac{2+\sqrt{2}}{4} \approx 0.854 \end{aligned} \tag{10.29}$$

This is more than 10% higher than can be achieved using a local hidden variable model for the CHSH game. Quantum mechanics thus predicts that something can be achieved – a success probability for the CHSH game, that no local theory of physics can possibly achieve.

This demonstrates conclusively that ***quantum mechanics cannot be replaced by a deeper theory that is local***.

10.2 Discussion

After Bell's theorem was discovered, there were two logical possibilities concerning nature: Maybe if we tried to implement the quantum mechanical strategy for the CHSH game we would only win 75% of the time, in accordance with what we expect to be able to achieve with local hidden variables. This would then constitute a theoretical prediction of quantum mechanics which is not verified in the lab, and would prove that quantum mechanics did not accurately describe physics. This would be a breakthrough.

On the other hand, we might indeed win approximately 85% of the time, in which case we would have an experimental observation which is inconsistent with the idea that our universe is described by a local theory of physics. This too, would be a breakthrough.

Bell's theorem was tested conclusively for the first time in 1972 by Clauser and Freedman. ***Their experiment confirmed the quantum mechanical prediction that nature is nonlocal.*** Since that initial experiment, it has been generalised and repeated and reproduced countless times, using every imaginable physical system, and every time they have shown resoundingly that quantum mechanics can achieve what a local theory of physics cannot.

The level of scepticism surrounding the results of these experiments has however been remarkable, with many explanations found for why the experiments did not actually carry out the thought experiment we analysed above faithfully enough, and finding ways that local hidden variable models could in fact utilise these discrepancies in fantastical ways to reproduce the observed results of the experiments. These have become known as ***loopholes***, and for real sceptics of quantum nonlocality, it was not clear whether if the real experiment would be carried out, whether quantum mechanics would finally be shown to fail.

However, in 2015 three groups independently performed experiments which were so careful, that everybody agrees that there were no loopholes left that could explain the findings using local hidden variables. We thus must accept that ***nature is fundamentally nonlocal***.

The study of ***quantum nonlocality*** is now a major branch of research within the quantum information community. Much like entanglement, which is one of the necessary quantum ingredients that leads to nonlocality, it is viewed as a resource that can be 'used' in order to achieve goals such as extremely reliable sources of random numbers, or quantum key distribution schemes which are provably secure against imperfections, to name just two.

On the fundamental side, there is still much that we would like to understand about nonlocality, including basic questions such as which states lead to nonlocality, and what is the precise relationship between entanglement and nonlocality? What are the limitations of quantum nonlocality? By trying to answer these questions, the hope is to learn about one of the most counter-intuitive phenomenon that can occur in nature.

11 Quantum Sensing

In this section we are going to introduce another interesting primitive task in quantum information science with promising technological applications, known as ***quantum sensing***. Very generally, in a sensing task a system will have some ***property*** which we would like to learn as ***efficiently as possible***. In quantum sensing, we use ***quantum probes*** in order to learn this property. Remarkably, we will see that ***entangled probes*** provide significant improvements in the ***precision*** with which we can learn properties of systems.

At an abstract level, quantum sensing isn't too dissimilar to tasks we have considerd previously in this course. It is closely related to the task of ***distinguishing quantum states***, which we considered in Sec. 2.2 and returned to in Sec. **??**. As we will see, we ***encode*** the property of interest into a state, and hence our ability to learn the property is intimately related to how well we can then distinguish states with different properties. It is also related to the Deutsch and Deutch-Jozsa problems considered in Sec. 4.6 and Sec. 4.7 respectively. There, we had a function f, from a set of functions, which was encoded into unitary transformations U_f. Our goal was to determine whether f had a certain property or not – i.e. to ***learn something*** about f. As we will see, sensing also naturally encodes the property of a system into a unitary.

The main difference that we will encounter here is that we are typically interested in a ***continuous parameter***, i.e. something which can vary continuously in some range, and our focus will be on how ***senstive*** we can be to ***small changes*** in this parameter, using quantum probes. This in particular allows us to assess our ***resolution***. Thus although our focus will be very different, it is useful to keep in mind that sensing is yet another instance of a general type of problem considered in quantum information science, concerning the way in which ***information can be stored and retrieved from quantum states***.

11.1 The sensing problem

Let us now introduce more carefully the problem we are interested in. We will present a concrete example to motivate the general setup, before keeping the presentation more abstract thereafter. Let's consider that within some region of space there is a uniform magnetic field, which we would like to determine as accurately as possible. We assume that we know the direction of the magnetic field (pointing vertically, in the $\hat{\mathbf{z}}$ direction), and we also know what the strength of the field ***should be***, but we want to be sensitive to ***small variations in the field*** around this base value.

The details of the physics of this situation aren't overly important, but so-called 'spin-1/2' particles (such as electrons) have two orthogonal states, which we can call $|0\rangle$ and $|1\rangle$. Assuming that we send such particles through the field, and that they spend time τ in the field, the dynamics, expressed in terms of the unitary evolution that this field will produce, is given by

$$U_{B,\tau} = e^{-iB\tau/\hbar}|0\rangle\langle 0| + e^{iB\tau/\hbar}|1\rangle\langle 1|, \tag{11.1}$$

which depends upon the field strength B and duration τ. We see that in fact the unitary depends only upon the product $B\tau$, and so to simplify, (and be a little more abstract), we will define a ***phase*** $\phi/2 = B\tau/\hbar$, so that the evolution is more succinctly given by

$$U_\phi = e^{-i\phi/2}|0\rangle\langle 0| + e^{i\phi/2}|1\rangle\langle 1|. \tag{11.2}$$

If the duration τ is known (which by assumption it is), then if we can determine ϕ, we can find the field strength $B = \hbar\phi/2\tau$ directly. Furthermore, as we said previously, we assume that the field should take a value, which we will denote B_0. This corresponds to a phase $\phi_0 = 2B_0\tau/\hbar$. If we are interested in ***variations*** in the field around B_0, this then corresponds to ***variations in the phase*** around ϕ_0.

We have thus managed to cast the problem ***solely in terms of the abstract phase*** ϕ, rather then in terms of the magnetic field and duration. It turns out that in other physical situations of interest for sensing, e.g. graviational wave sensing, birefrigence sensing in optical fibres, and many other examples, we can ***always arrive back at exactly this abstract formulation of the problem***, in terms of a ***varying phase*** ϕ. For this reason, we will now focus on this abstract form of the problem, to gain a general understanding.

11.2 Single-probe sensing

In order to build some intuition, and formalise our figure of merit – how ***sensitive*** we are to variations in ϕ – it will be instructive to consider the simplest possible situation, of using a single qubit as a probe. To that end, consider that we prepare as our probe in the state $|+\rangle = (|0\rangle + |1\rangle)/\sqrt{2}$. Probing the magnetic field corresponds to applying the unitary U_ϕ, in which case, our state becomes

$$\begin{aligned} |\psi_\phi\rangle &= U_\phi|+\rangle, \\ &= (e^{-i\phi/2}|0\rangle\langle 0| + e^{i\phi/2}|1\rangle\langle 1|)\frac{1}{\sqrt{2}}(|0\rangle + |1\rangle), \\ &= \frac{1}{\sqrt{2}}(e^{-i\phi/2}|0\rangle + e^{i\phi/2}|1\rangle). \end{aligned} \tag{11.3}$$

We can now perform a measurement on this state in order to try and learn something about ϕ. A good choice is to measure $X = |+\rangle\langle +| - |-\rangle\langle -|$. In this case, the probabilities are

$$\begin{aligned} \text{Prob}(+1) &= \||+\rangle\langle +|\psi_\phi\rangle\|^2, & \text{Prob}(-1) &= \||-\rangle\langle -|\psi_\phi\rangle\|^2, \\ &= \left|\frac{e^{-i\phi/2} + e^{i\phi/2}}{2}\right|^2, & &= \left|\frac{e^{-i\phi/2} - e^{i\phi/2}}{2}\right|^2, \\ &= \cos^2\frac{\phi}{2}, & &= \sin^2\frac{\phi}{2}, \end{aligned} \tag{11.4}$$

and the expectation value is

$$\begin{aligned} \langle X\rangle_\phi &= \cos^2\frac{\phi}{2} - \sin^2\frac{\phi}{2}, \\ &= \cos\phi. \end{aligned} \tag{11.5}$$

This means that if we were to repeat this procedure many many times, and collect statistics, we could estimate the expectation value $\langle X\rangle_\phi$, from which we could then infer the phase by $\phi = \cos^{-1}(\langle X\rangle_\phi)$.

The interesting question is what is our ***resolution***, that is, how accurately can we hope to determine ϕ. This will depend on how ***sensitive*** we are to variations in ϕ. Where would a limitation on our sensitivity or resolution arise from? The answer is that we will ***always have some uncertainty in*** $\langle X\rangle_\phi$ ***arising from the finite statistics we are able to collect***. We therefore need to understand how this translates into uncertainty in ϕ. In order to ***resolve*** two distinct values of ϕ, we need them to be separated by ***more*** than the uncertainty we have in ϕ. This is the origin of our resolution limit. This idea is succinctly captured in Fig. 8.

In particular, there we can see that how rapidly $\langle X\rangle_\phi$ varies with ϕ determines the uncertainty Δ_ϕ we have in ϕ. Formally, assuming that ϕ takes on the particular value $\phi = \phi_0$, if the estimated expectation value is x_0 with uncertainty Δx_0, then the uncertainty $\Delta\phi_0$ in the value $\phi_0 = \cos^{-1}(x_0)$ is

$$\Delta\phi_0 = \frac{\Delta x_0}{\left|\frac{d\langle X\rangle_\phi}{d\phi}\Big|_{\phi_0}\right|}, \tag{11.6}$$

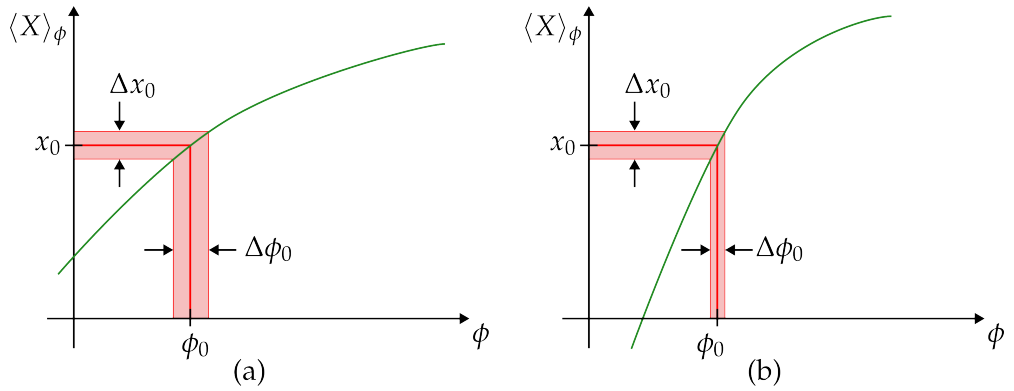

Figure 8: This figure depicts a general case, where we have a phase ϕ we are trying to learn, using some quantum probe state, and performing some measurement (i.e. this is not specific to the example given, but it is to depict the general case). We denote the expectation value of the performed measurement by $\langle X\rangle_\phi$, and assume that after collecting statistics, we obtain an estimate of this, denoted x_0, with uncertainty Δx_0. From the graph, we can read off the associated value of ϕ_0, and its residual uncertainty $\Delta\phi_0$. In cases (a) and (b) we consider two hypothetical situations, in the first where $\langle M\rangle_\phi$ is ***less sensitive*** to changes to ϕ (i.e. corresponding to a shallower green curve), and in the second where $\langle M\rangle_\phi$ is ***more sensitive*** to changes to ϕ (i.e. corresponding to a steeper green curve). We see that for the same uncertainty Δx_0, in the latter case we have much less residual uncertainty $\Delta\phi_0$, ***due to the sensitivity***. We can reliably distinguish ϕ_0 from any value of ϕ outside the pink shaded region on the ϕ axis. The width of this region can therefore be considered our ***resolution***. In case (b) we therefore see that we have a higher resolution than case (a), due to the increased sensitivity.

i.e. it is precisely the uncertainty in x_0, rescaled by the rate of change of the expectation value at the estimated point. This is a standard result from statistics (known as propagation of uncertainty/error), and follows by taking a Taylor expansion of $\langle X\rangle_\phi$ about the point ϕ_0.

The final question is what determines Δx_0? As we said above, this arises due to finite statistics, but more formally, we know that it is the ***standard deviation*** that sets the overall scale of the uncertainty. In particular, in (11.4) we calculated the probabilities for the measurement results $+1$ and -1, and the associated expectation value $\langle X\rangle_\phi$. The second moment of this distribution is

$$\begin{aligned}\langle X^2\rangle_\phi &= (+1)^2\times\text{Prob}(+1)+(-1)^2\times\text{Prob}(-1),\\ &= \cos^2\frac{\phi}{2}+\sin^2\frac{\phi}{2},\\ &= 1,\end{aligned}\tag{11.7}$$

and so the standard deviation is

$$\begin{aligned}\Delta X_\phi &= \sqrt{\langle X^2\rangle-\langle X\rangle^2},\\ &= \sqrt{1-\cos^2\phi},\\ &= |\sin\phi|.\end{aligned}\tag{11.8}$$

After N repetitions of the procedure, assuming the specific phase $\phi=\phi_0$, the uncertainty Δx_0 we have in our observed estimate x_0 of the (actual) expectation value $\langle X\rangle_{\phi_0}$ is

$$\Delta x_0 = \frac{\Delta X_{\phi_0}}{\sqrt{N}}.\tag{11.9}$$

That is, our uncertainty decreases as $1/\sqrt{N}$, with the standard deviation ΔX setting the overall scale. This is once again a standard result from statistics, known as the "***standard error in the mean***".

Putting everything together, in this example of single-probe sensing, we have $\frac{d\langle X\rangle_\phi}{d\phi} = -\sin\phi$, and so the resolution $\Delta\phi_0$ around the point ϕ_0 is

$$\Delta\phi_0 = \frac{\frac{|\sin\phi_0|}{\sqrt{N}}}{|-\sin\phi_0|} = \frac{1}{\sqrt{N}}. \tag{11.10}$$

This shows us that in this case, as is intuitively clear, in order to increase resolution, we must send through more probe states. More quantitatively, it shows us that, for example, to double the resolution (decrease the uncertainty by a factor of 2), we in fact need to send through four times as many states, while to increase the resolution by a factor of 10, we would need to use 100 times as many probes! Thus while it is possible to increase the resolution by repetition, we see it is rather ***expensive***. The main achievement of quantum sensing – as we will now see – is that ***there are much more efficient quantum probes***, in fact, ones where in order to increase the resolution by a factor of n only requires us to increase the number of probes by the same factor of n – a quadratic (squared) improvement over the above example.

11.3 n-probe sensing

The main idea now is to consider probing the unitary U_ϕ n ***times in parallel***. That is, in a more physical language, instead of sending particles through a magnetic field one at a time, we will consider sending through n particles at once – potentially in an interesting ***entangled*** state – and then measure those particles in order to learn about the field. As in the above, we will still be interested in being as sensitive as possible to small changes in the field.

The simplest n-probe state that we could consider would be to send n identical qubits in the product state considered above, i.e. to use the state $|+\rangle^{\otimes n}$. It will be useful to calculate what happens here, as it will act as our ***benchmark*** for the more interesting entangled probe strategy that will follow. In the present case, the state after applying the unitary U_ϕ ***on each copy*** becomes

$$\begin{aligned} |\psi_\phi^n\rangle &= \left(U_\phi \otimes \cdots \otimes U_\phi\right)|+\rangle|+\rangle\cdots|+\rangle, \\ &= \left(\frac{e^{-i\phi/2}|0\rangle + e^{i\phi/2}|1\rangle}{\sqrt{2}}\right)^{\otimes n}. \end{aligned} \tag{11.11}$$

We now measure X on each qubit, and it is natural to ***add*** the outcomes together, meaning we will obtain results ranging from $+n$ to $-n$ in integer steps. The resulting probability distribution is a ***binomial*** distribution, with

$$\text{Prob}(k) = \binom{n}{\frac{n+k}{2}} \left(\cos^2\frac{\phi}{2}\right)^{\frac{n+k}{2}} \left(\sin^2\frac{\phi}{2}\right)^{\frac{n-k}{2}}, \tag{11.12}$$

where $\binom{n}{\ell} = \frac{n!}{\ell!(n-\ell)!}$ is the binomial coefficient. The average and standard deviation of this distribution are

$$\langle X\rangle_\phi = n\cos\phi, \qquad \Delta X_\phi = \sqrt{n}|\sin\phi|, \tag{11.13}$$

which are, respectively n times the average (11.5) and $\sqrt{n}$ times the standard deviation (11.8) from the single-probe case, as they should be, since this simple n-probe procedure is nothing but the combination (summation) of n single-probe procedures.

Thus, if we consider the specific phase $\phi = \phi_0$ again, after N repetitions of the procedure (each involving an n-qubit probe state now, so that the ***total*** number of qubits is nN), then from (11.9)

we would obtain an estimate x_0, with statistical uncertainty

$$\Delta x_0 = \frac{\Delta X_{\phi_0}}{\sqrt{N}} = \sqrt{\frac{n}{N}}|\sin\phi_0|. \tag{11.14}$$

Combining this with the fact that $\frac{d\langle X\rangle_\phi}{d\phi} = -n\sin\phi$, we see that, from (11.6), the resolution becomes

$$\Delta\phi_0 = \frac{\Delta x_0}{\left|\frac{d\langle X\rangle_\phi}{d\phi}\Big|_{\phi_0}\right|} = \frac{1}{\sqrt{nN}}, \tag{11.15}$$

i.e. we obtain ***exactly the same result*** as (11.10), except we have nN instead of N (the total number of qubits used when repeating the n-probe procedure N times). What this confirms is the intuitive result that by averaging before repeating, we don't gain any advantage over just repeating. This result – that using n unentangled probe states in parallel leads to a precision scaling as $1/\sqrt{n}$ is referred to as the '***standard quantum limit***'.

This has set the stage for now considering the much more interesting case, of using an n-qubit entangled probe. We will see that we can surpass the standard quantum limit on precision, and achieve a precision known as the '***Heisenberg limit***'.

Let us now imagine that before probing the phase ϕ, we first ***entangle*** all the qubits, so that they end up in the following interesting state:

$$|\omega^n\rangle = \frac{|0\rangle^{\otimes n} + |1\rangle^{\otimes n}}{\sqrt{2}}, \tag{11.16}$$

This state is known as the 'GHZ' state (after Greenberger, Horne and Zeilinger). When $n = 2$, this is just the maximally entangled Bell state we have seen many times before. This state is one of the most interesting states of multiple qubits. How can we prepare this state? We can use the CNOT operation repeatedly from (3.9). In particular, consider preparing $|+\rangle|0\rangle\cdots|0\rangle$, and applying CNOTs sequentially between qubits, starting with the first and second, then the second and third, and so forth, then we see that

$$|+\rangle|0\rangle^{\otimes n-1} \to \frac{|0\rangle|0\rangle + |1\rangle|1\rangle}{\sqrt{2}}|0\rangle^{\otimes n-2} \to \frac{|0\rangle|0\rangle|0\rangle + |1\rangle|1\rangle|1\rangle}{\sqrt{2}}|0\rangle^{\otimes n-3} \to \cdots \to |\omega^n\rangle, \tag{11.17}$$

i.e. each CNOT 'grows' a GHZ state from $|\omega^k\rangle$ to $|\omega^{k+1}\rangle$.

If we now consider probing U_ϕ using this GHZ state, we arrive at

$$\begin{aligned} |\omega_\phi^n\rangle &= (U_\phi \otimes \cdots U_\phi)|\omega_n\rangle, \\ &= \frac{e^{-in\phi/2}|0\rangle^{\otimes n} + e^{in\phi/2}|1\rangle^{\otimes n}}{\sqrt{2}}. \end{aligned} \tag{11.18}$$

This is somewhat similar to the single-probe state (11.3), except we have replaced $|0\rangle$ and $|1\rangle$ by $|0\rangle^{\otimes n}$ and $|1\rangle^{\otimes n}$ respectively, and the phase ϕ has become $n\phi$. This already hints at a nice type of ***amplification***, as $|0\rangle^{\otimes n}$ and $|1\rangle^{\otimes n}$ are still two orthogonal states, and yet now ϕ is n times larger.

We could consider measuring all n qubits, but in fact we can do something much smarter. If we consider applying the preparation procedure for the GHZ state in reverse to the state $|\omega_\phi^n\rangle$, that is, we apply CNOTs sequentially, starting with the final two qubits, and working back all the way to the first two, we find that

$$\begin{aligned} |\omega_\phi^n\rangle &\to \frac{e^{-in\phi/2}|0\rangle^{\otimes n-1} + e^{in\phi/2}|1\rangle^{\otimes n-1}}{\sqrt{2}}|0\rangle \to \frac{e^{-in\phi/2}|0\rangle^{\otimes n-2} + e^{in\phi/2}|1\rangle^{\otimes n-2}}{\sqrt{2}}|0\rangle^{\otimes 2} \\ &\to \cdots \to \frac{e^{-in\phi/2}|0\rangle + e^{in\phi/2}|1\rangle}{\sqrt{2}}|0\rangle^{\otimes n-1}. \end{aligned} \tag{11.19}$$

That is, we return $n-1$ qubits back to the state $|0\rangle$, and the remaining qubit ends up in the single-probe state $|\psi_{n\phi}\rangle = \frac{1}{\sqrt{2}}(e^{-in\phi/2}|0\rangle + e^{in\phi/2}|1\rangle)$ of the form (11.3), with the phase genuinely amplified now. We can therefore proceed ***identically*** to the single probe case: we measure X on this qubit, and all of the analysis from Section 11.2 holds, except we now replace ϕ by $n\phi$. In particular, this means that here

$$\langle X\rangle_\phi = \cos n\phi, \qquad \Delta X_\phi = |\sin n\phi|. \tag{11.20}$$

Crucially, now we see that our sensitivity to changes in ϕ is increased due to this amplification, in particular

$$\frac{d\langle X\rangle_\phi}{d\phi} = -n\sin n\phi, \tag{11.21}$$

which is crucially now n times larger than it was previously. Putting everything together, our resolution, assuming that ϕ takes on the specific value $\phi = \phi_0$, and that we repeat this GHZ probing strategy N times to collect statistics as always, is now

$$\Delta\phi_0 = \frac{\Delta x_0}{\left|\frac{d\langle X\rangle_\phi}{d\phi}\Big|_{\phi_0}\right|} = \frac{\frac{|\sin n\phi_0|}{\sqrt{N}}}{|-n\sin n\phi_0|} = \frac{1}{n\sqrt{N}} \tag{11.22}$$

Remarkably, comparing to (11.15), we see that our scaling in n has improved significantly: whereas previously our precision scaled as $1/\sqrt{n}$ (which we referred to as the standard quantum limit), it now scales as $1/n$. ***Our entangled strategy is far out-performing the product strategy above***. This improved scaling in the precision of our estimate with n – now as $1/n$ – is known as the '***Heisenberg limit***' of quantum sensing, and can be shown to be the ***best possible scaling that can be achieved***.

What have we learnt? We now see that quantum mechanics provides a very efficient way to improve sensitivity if we use entanglement: whereas without it, if we want to double our precision relative to some baseline (i.e. we assume that the number of repetitions N used to collect statistics is fixed), we had to increase the number of probe states we sent through in parallel by a factor of 4. Now we can achieve the same improvement, by only sending through 2 probes in parallel. The price we pay is that we have to entangle and disentangle our probes before and after sensing, and this may indeed be tricky, but assuming we can do so, we become much more sensitive to small changes in parameter than when we don't use entanglement.

We could alternatively phrase things the other way: if we consider the total number of probe qubits k to be fixed, if we don't entangle any of them, then our precision is $1/\sqrt{k}$, while if we entangle them into n-qubit GHZ states (and so repeat the procedure $N = k/n$ times), then our precision is $1/\sqrt{nk}$, an improvement by the factor $1/\sqrt{n}$.

It is not only the GHZ state that leads to the Heisenberg limit. In the context of optics, a closely related state is known as the '$N00N$ state', which achieves the same precision, as well as other states. However, the analysis here sets out the general framework, which can then be applied to any other choice of probe state.

12 The Quantum Internet

As a final section of this course we will consider some of the basic aspects of quantum information science that will underly a potential future ***quantum internet***. What is the quantum internet? Well, much like the (classical) internet, the basic idea is that one day – potentially in the near future – there will be a global scale network of nodes, able to exchange ***quantum communication*** and ***quantum entanglement***. There are numerous reasons that are already clear for why this might be desirable, but just as it wasn't entirely clear what the internet would evolve into when it was first conceived, it is also pretty safe to assume that the same will be true of a potential quantum internet.

Let's start with some of the topics that we have already considered in this course. We have seen that if two parties share ***quantum entanglement***, then via the process of ***quantum teleportation*** it is possible to send ***quantum states*** using only ***classical communication***. On the one hand, this shows that if we are able to ***distribute entanglement*** around the globe, then this could be used to facilitate ***quantum communication***. On the other hand, this is furthermore a desirable way to communicate quantum information, as in reality ***quantum communication will always be noisy***, whereas classical communication can be thought of as being effectively noise-free, given our ability to, e.g. repeat messages to overcome any noise in the communication channel, or use other ***error-correction*** techniques.

What might be the use of such quantum information? There are many possibilities that could be mentioned. First, it is conceivable that the first ***quantum computers*** will be in the ***cloud*** (due to their size and expense, at the very least). It is then imaginable that users might want to send jobs to these machines, and might want to keep some level of privacy. Being able to send qubits to a quantum computer, encoding data, and receiving qubits back is therefore a distinct functionality that might arise.

The second major use will undoubtedly be in cryptographic situations. We have already seen how quantum mechanics allows for ***quantum key distribution (QKD)***: allowing for users to generate shared secret keys, which can then be used to encrypt classical data, for transmission over the internet. Quantum communication will facilitate the use of quantum key distribution. In fact, modern schemes for QKD are ***entanglement*** and even ***non-locality*** based, and so the distribution of entanglement will likely also be important for QKD.

Key distribution is however only the simplest cryptographic task that one might want to use quantum mechanics for, and there are a wealth of other ***cryptographic primitives*** which a quantum internet might make broadly available. These include ideas such as ***digital signatures*** (i.e. verfication), as well as ***secret sharing*** (i.e. making agreements with untrusted parties), as well as in ***voting***. This is a hugely active area of research, and one of the most advanced technologically.

Finally, we will mention that ***synchronisation*** is a crucial functionality on the internet, where it is envisaged that a quantum internet could become indispensable.

As we said above, at the core of the quantum internet will be ***entanglement distribution***, which is the primitive task we will focus on in this final section. In particular, there are two challenges that we will address, each of which allow us to delve into a fundamental task in quantum information. The first of these is the need to distribute entanglement over ***long distances***. The problem with this is that, realistically, just sending qubits over available communication channels (e.g. fibre optics, or even through the air) is infeasible at the global distances required, as the qubits essentially get lost on the way (and since they can't be cloned, this appears to be a fundamental problem rather than a technological one!). Remarkably, there is in fact a way to ***entangle systems at a distance!*** This is called ***entanglement swapping***, and can be seen as a generalisation of ***teleportation***. Using entanglement swapping, we can distribute entanglement over very long distances using the idea of a so-called ***entanglement repeater***.

Second, in order to be useful for all of the exciting tasks we briefly mentioned above, the entanglement shared over the quantum internet needs to be ***of very high quality***. If it is low quality

entanglement, then very quickly it becomes useless. In reality, we never distribute the perfect maximally entangled states we want to distribute, and again it appears like this might be a fundamental roadblock to any realistic quantum internet. Also remarkably, we can overcome this, using a primitive known as ***entanglement purification***: this allows us to turn a ***large number of low qualtiy entangled states into a smaller number of high-quality states***.

In what follows we will outline the basics of these two topics, which together lay the foundations for the quantum internet.

12.1 Entanglement swapping and quantum repeaters

Consider the following simple question: What happens if Alice tries to teleport to Bob half of an entangled state? More precisely, imagine that Alice shares the entangled state

$$|\Phi\rangle = \alpha|0\rangle|0\rangle + \beta|1\rangle|1\rangle, \tag{12.1}$$

with Charlie, such that Alice holds one qubit, and Charlie holds the other (and we assume $\alpha \neq 0$ or 1). Alice now decides that she would like to teleport her qubit to Bob (with whom she also shares the maximally entangled state, as in the standard teleportation protocol). What will happen? There are a number of initial reactions one might have. For example, you might imagine that Alice will teleport her reduced density operator ρ_A to Bob? You might imagine that it only works if Alice is unentangled from Charlie?

It turns out that ***teleportation works just as before, and maintains the entanglement with Charlie!*** That is, Alice really teleports her qubit, and this qubit retains the entanglement it has with Charlie's qubit, so that at the end of the protocol, Bob and Charlie's qubits become entangled. This is remarkable since ***Bob and Charlie's qubits may never have interacted ever, and may never have even been in the same place before***. Nevertheless, we are able to entangle them ***at a distance***, using teleportation.

To see this, we can start in a similar fashion to how we started with teleportation. The initial state, with the qubits labelled for convenience, is

$$|\Psi\rangle = (\alpha|0\rangle_C|0\rangle_A + \beta|1\rangle_C|1\rangle_A)\frac{|0\rangle_{A'}|0\rangle_B + |1\rangle_{A'}|1\rangle_B}{\sqrt{2}}, \tag{12.2}$$

where we have labelled the two qubits held by Alice with A and A' respectively. It will be useful to do two things at this stage: first, the ordering of the qubits will make it difficult to see what is going on, so we will need to write them in a different order. ***We are not making any kind of physical change here, we are just choosing the write the state in a more convenient form*** (indeed it was our choice to write the qubits in the order Charlie-Alice-Alice'-Bob, and we are free to choose any other order we like to represent the state on paper). It will be convenient to use the order Alice-Alice'-Bob-Charlie. In this ordering, the state is written

$$\begin{aligned}|\Psi\rangle = \frac{\alpha}{\sqrt{2}}|0\rangle_A|0\rangle_{A'}|0\rangle_B|0\rangle_C + \frac{\alpha}{\sqrt{2}}|0\rangle_A|1\rangle_{A'}|1\rangle_B|0\rangle_C \\ + \frac{\beta}{\sqrt{2}}|1\rangle_A|0\rangle_{A'}|0\rangle_B|1\rangle_C + \frac{\beta}{\sqrt{2}}|1\rangle_A|1\rangle_{A'}|1\rangle_B|1\rangle_C.\end{aligned} \tag{12.3}$$

Second, similar to how we analysed teleportation, it is useful to now rewrite Alice's two qubits in the Bell basis, since she will perform a measurement in this basis, just as in teleportation. In particular, from the Bell basis,

$$\begin{aligned}|\Phi_I\rangle &= \frac{1}{\sqrt{2}}(|0\rangle|0\rangle + |1\rangle|1\rangle), & |\Phi_X\rangle &= \frac{1}{\sqrt{2}}(|0\rangle|1\rangle + |1\rangle|0\rangle), \\ |\Phi_Y\rangle &= \frac{1}{\sqrt{2}}(|0\rangle|1\rangle - |1\rangle|0\rangle), & |\Phi_Z\rangle &= \frac{1}{\sqrt{2}}(|0\rangle|0\rangle - |1\rangle|1\rangle).\end{aligned} \tag{12.4}$$

we immediately see that

$$|0\rangle|0\rangle = \frac{|\Phi_I\rangle + |\Phi_Z\rangle}{\sqrt{2}}, \qquad |0\rangle|1\rangle = \frac{|\Phi_X\rangle + |\Phi_Y\rangle}{\sqrt{2}},$$
$$|1\rangle|0\rangle = \frac{|\Phi_X\rangle - |\Phi_Y\rangle}{\sqrt{2}}, \qquad |1\rangle|1\rangle = \frac{|\Phi_I\rangle - |\Phi_Z\rangle}{\sqrt{2}}. \tag{12.5}$$

Substituting these into (12.3) for Alice's two qubits, we see that

$$\begin{aligned}|\Psi\rangle = &\frac{\alpha}{2}(|\Phi_I\rangle_{AA'} + |\Phi_Z\rangle_{AA'})|0\rangle_B|0\rangle_C + \frac{\alpha}{2}(|\Phi_X\rangle_{AA'} + |\Phi_Y\rangle_{AA'})|1\rangle_B|0\rangle_C \\ &+ \frac{\beta}{2}(|\Phi_X\rangle_{AA'} - |\Phi_Y\rangle_{AA'})|0\rangle_B|1\rangle_C + \frac{\beta}{2}(|\Phi_I\rangle_{AA'} - |\Phi_Z\rangle_{AA'})|1\rangle_B|1\rangle_C.\end{aligned} \tag{12.6}$$

Collecting together the states in Alice's qubits, this is further equal to

$$\begin{aligned}|\Psi\rangle = &\frac{1}{2}|\Phi_I\rangle_{AA'}(\alpha|0\rangle_B|0\rangle_C + \beta|1\rangle_B|1\rangle_C) + \frac{1}{2}|\Phi_X\rangle_{AA'}(\alpha|1\rangle_B|0\rangle_C + \beta|0\rangle_B|1\rangle_C) \\ &+ \frac{1}{2}|\Phi_Y\rangle_{AA'}(\alpha|1\rangle_B|0\rangle_C - \beta|0\rangle_B|1\rangle_C) + \frac{1}{2}|\Phi_Z\rangle_{AA'}(\alpha|0\rangle_B|0\rangle_C - \beta|1\rangle_B|1\rangle_C),\end{aligned} \tag{12.7}$$

Interestingly, we see that written this way, the state is a superposition, of Alice's qubits being in one of the four Bell states, and correspondingly the qubits of Bob and Charlie being in one of four states, all closely related to the state shared by Alice and Charlie.

As in teleportation, Alice now measures an operator which has the four Bell states $|\Phi_I\rangle_{AA'}$, $|\Phi_X\rangle_{AA'}$, $|\Phi_Y\rangle_{AA'}$ and $|\Phi_Z\rangle_{AA'}$ as eigenstates, with corresponding eigenvalues λ_I, λ_X, λ_Y and λ_Z respectively. Given the form of the state, we can read off the result. With probability $\text{Prob}(\lambda_I) = |\frac{1}{2}|^2 = \frac{1}{4}$ Alice obtains the result λ_I, and the state after measurement becomes

$$|\Psi_I\rangle = |\Phi_I\rangle_{AA'}(\alpha|0\rangle_B|0\rangle_C + \beta|1\rangle_B|1\rangle_C), \tag{12.8}$$

That is, Bob and Charlie end up in the ***entangled state*** $|\Phi\rangle$ originally shared by Alice and Charlie. We note also that Alice's two qubits now end up entangled with each other, in the state $|\Phi_I\rangle$, and hence Alice has destroyed her entanglement with both Bob and Charlie.

Similarly, Alice obtains the outcome λ_X with probability $\text{Prob}(\lambda_X) = |\frac{1}{2}|^2 = \frac{1}{4}$, in which case the state after measurement becomes

$$|\Psi_X\rangle = |\Phi_X\rangle_{AA'}(\alpha|1\rangle_B|0\rangle_C + \beta|0\rangle_B|1\rangle_C). \tag{12.9}$$

Bob and Charlie are not in the correct state $|\Phi\rangle$, however, if Bob were to apply the same correction as in teleportation to his qubit – i.e. to act with X on his qubit, then the state of Bob and Charlie would become

$$(X \otimes I)(\alpha|1\rangle_B|0\rangle_C + \beta|0\rangle_B|1\rangle_C) = (\alpha|0\rangle_B|0\rangle_C + \beta|1\rangle_B|1\rangle_C) = |\Phi\rangle, \tag{12.10}$$

again the state originally shared by Alice and Charlie.

We leave it as an exercise to show that in the other two cases, when Alice obtains either the outcome λ_Y or λ_Z, then by Bob performing ***the same correction as in teleportation***, Bob and Charlie end up in the state $|\Phi\rangle$.

In summary, what this shows is that if Alice teleports a qubit which is entangled to another qubit (in this case, her qubit in the state $|\Phi\rangle$), then by following the standard teleportation protocol, of measuring her qubits with an operator that has the four Bell states as eigenstates and then communicating the result to Bob (corresponding to two classical bits of communication), who then corrects the state by applying the appropriate Pauli operator, then Alice and Bob manage to ***swap entanglement***, so that Bob and Charlie end up in the exact same entangled state initially shared by Alice and Charlie!

What is particularly remarkable about this is that the qubits of Bob and Charlie ***never interact*** – we may imagine that Charlie prepared the state $|\Phi\rangle$ (by interacting two qubits together), before sending one to Alice. We may similarly assume that Bob prepared the state $|\Phi_I\rangle$ in a similar way, before sending a qubit to Alice. The qubits left in Bob's and Charlie's possession were therefore ***never*** in the same place at the same time. Nevertheless, Alice is able to ***entangle them at a distance***, using the teleportation protocol and classical communication.

How is this useful for distributing entanglement over long distances? Well, assume that Charlie, Alice and Bob are all in a line, with Charlie 10km to the west of Alice, and Bob 10km to the east. Bob and Charlie are therefore 20km from each other. Assume also that if we try to send a qubit further than 10km we are guaranteed that it will get lost. This means we cannot directly send a qubit from Charlie to Bob, since they are too far from each other. Using entanglement swapping however, we can first distribute entanglement over the shorter distances of 10km, and then use the protocol to convert this into entanglement over the 20km distance. This functionality is called a ***quantum entanglement repeater***. By iterating this process, we can in principle distribute entanglement over ***arbitrarily large distances***.

You still may not be convinced of the utility of this however. Surely we could just send a qubit from Charlie to Alice, who could then forward it to Bob, and achieve the same result? That is true, but one big advantage here is that we can ***establish the entanglement ahead of time***, whereas in the above, we can only communicate ***after we know the quantum message we want to send***, and in this case it might fail, requiring us to restart the whole protocol over 20km. In the entanglement swapping version, the entanglement can be distributed ahead of time, and we can also use techniques to be ***certain*** that we have succeeded in doing so.

This is similar for entangled-based QKD protocols, where the communication is not about sending a message, but about sharing entanglement at a distance. Entanglement swapping allows us to establish that entanglement in advance, in order to then use for QKD, or any other application we have in mind.

12.2 Entanglement purification

In the previous subsection we saw how we could overcome the challenge of distributing entanglement over large distances, using entanglement swapping to produce so-called quantum repeaters. In that subsection we however assumed that the entangled states initially shared by the parties are ***perfect***. This isn't realistic, and hence the second major challenge facing the quantum internet is how to ***overcome noise***. In this subsection, we will show that we can in fact achieve something beautiful, known as ***entanglement purification***.

To explain the basic idea, let us consider that in the process of creating and distributing an entangled quantum state certain errors can occur. In particular, let's assume that there is some chance that a qubit flips its state from $|0\rangle$ to $|1\rangle$ at random. More precisely, with probability p we assume that an X operator is applied to each qubit in the state $|\Phi_I\rangle$ (and with probability $(1-p)$ nothing happens). That means with probability $(1-p)^2$ the state remains $|\Phi_I\rangle$. With probability $(1-p)p$ only the second qubit flips, in which case the state becomes $|\Phi_X\rangle = (I \otimes X)|\Phi_I\rangle$. With the same probability $p(1-p)$ only the first qubit flips, and we end up again in the state $|\Phi_X\rangle = (X \otimes I)|\Phi_I\rangle$, and with probability p^2 both qubits flip, but in this case the state in fact remains the same, since $(X \otimes X)|\Phi_I\rangle = |\Phi_I\rangle$ (all of which can be easily verified by direct calculation). We see that, due to this random process, we end up with the ***ensemble***

$$\{|\Phi_I\rangle, (1-p)^2 + p^2; |\Phi_X\rangle, 2p(1-p)\}. \tag{12.11}$$

That is, we end up in either the state $|\Phi_I\rangle$ with total probability $q = p^2 + (1-p)^2$ or the state $|\Phi_X\rangle$ with probability $1 - q = 2p(1-p)$. As we saw in Section 5, we should describe this ensemble by a ***density operator***

$$\rho = q|\Phi_I\rangle\langle\Phi_I| + (1-q)|\Phi_X\rangle\langle\Phi_X|. \tag{12.12}$$

This density operator is a more accurate description of the realistic state that Alice and Bob would share after distributing entanglement between themselves. If $q = 1$ we recover the case we previously considered (the state is just $|\Phi_I\rangle$). We are thus primarily interested in cases where $q \approx 1$, modelling the case of a small chance of the qubits having been flipped in the process of distribution. The goal of Alice and Bob is to ***improve*** their state somehow, to make it closer to the ideal case of $q = 1$.

In order to achieve this, let us now assume that Alice and Bob have created and distributed ***two copies*** of this state, and so share $\rho \otimes \rho$, with each of them holding two qubits. We will let them interact locally with their respective qubits (e.g. applying unitaries, and performing measurements) and to ***communicate classical information***. Both of these can be thought of as 'cheap' things to do – they certainly don't require any entanglement between Alice and Bob, and this is what we are treating as being the precious resource. After they have finished, we would like them to have produced a single pair of qubits in an entangled state, which crucially is ***more highly entangled*** than ρ. Additionally, we will only ask that they achieve this ***probabilistically***: that is, sometimes we will allow Alice and Bob to fail. In this case, we can imagine that they simply discard their qubits, and start all over again. Assuming that their probability of failure isn't too high, after repeating this procedure a few times, they are overwhelmingly likely to succeed (that is, they would have to be ***very*** unlucky to fail every single time they try).

The procedure they follow is surprisingly simple:

1. Alice and Bob both individually apply the CNOT unitary (3.9) to their pair of qubits.
2. They each measure Z on their second qubit, obtaining the result $+1$ or -1.
3. They communicate classically, telling each other the result of their Z measurement. If they obtain the ***same*** result then they keep their first qubit; if they obtain ***different*** results they have failed, and restart the whole procedure.

We now want to analyse this procedure, and see that, if they succeed in obtaining the same measurement results, that their remaining pair of qubits are indeed improved, and are more entangled. In order to carry out this analysis, it will be insightful to consider individually the four distinct global states that they could share, i.e. to consider the four states in the ensemble separately, as will become clear in what follows.

We will start of therefore by considering what happens if they share the state $|\Phi_I\rangle|\Phi_I\rangle$, which they share with probability q^2. After step 1, the state shared becomes

$$\begin{aligned} |\Phi_I\rangle|\Phi_I\rangle &= \frac{1}{2}\left(|0\rangle|0\rangle|0\rangle|0\rangle + |0\rangle|0\rangle|1\rangle|1\rangle + |1\rangle|1\rangle|0\rangle|0\rangle + |1\rangle|1\rangle|1\rangle|1\rangle\right), \\ &\xrightarrow{CNOT\otimes CNOT} \frac{1}{2}\left(|0\rangle|0\rangle|0\rangle|0\rangle + |0\rangle|0\rangle|1\rangle|1\rangle + |1\rangle|1\rangle|1\rangle|1\rangle + |1\rangle|1\rangle|0\rangle|0\rangle\right), \\ &= |\Phi_I\rangle|\Phi_I\rangle, \end{aligned} \tag{12.13}$$

(where we recall that the CNOTs act between the first and third, and second and fourth qubits respectively, since these are the pairs of qubits held by Alice and Bob). That is, the state is unchanged in step 1. If Alice and Bob now each measure their second qubit, as we have seen many times before, they will ***always obtain the same outcome***: half of the time they will both obtain the outcome $+1$, and half of the time they will obtain the outcome -1. Therefore, when they communicate, they will never fail, and so they always keep their first pair of qubits, which are in the state $|\Phi_I\rangle$.

We can now move onto the case where they share the state $|\Phi_I\rangle|\Phi_X\rangle$, which they share with

probability $q(1-q)$. We can proceed as before, to see that after step 1,

$$\begin{aligned}|\Phi_I\rangle|\Phi_X\rangle &= \frac{1}{2}\left(|0\rangle|0\rangle|0\rangle|1\rangle + |0\rangle|0\rangle|1\rangle|0\rangle + |1\rangle|1\rangle|0\rangle|1\rangle + |1\rangle|1\rangle|1\rangle|0\rangle\right), \\ &\xrightarrow{CNOT\otimes CNOT} \frac{1}{2}\left(|0\rangle|0\rangle|0\rangle|1\rangle + |0\rangle|0\rangle|1\rangle|0\rangle + |1\rangle|1\rangle|1\rangle|0\rangle + |1\rangle|1\rangle|0\rangle|1\rangle\right), \\ &= |\Phi_I\rangle|\Phi_X\rangle. \end{aligned} \tag{12.14}$$

In this case, the CNOTs also leave the state unchanged. Interestingly, now the second pair of qubits are in the state $|\Phi_X\rangle$, and in step 2, this has the property that Alice and Bob will also obtain ***different*** measurement results. Thus, when they communicate in step 3, they will fail all of the time, and always discard their remaining qubits, even though they were in the ideal state $|\Phi_I\rangle$.

The next case to consider is when they share $|\Phi_X\rangle|\Phi_I\rangle$, which they share with probability $(1-q)q$. We find now

$$\begin{aligned}|\Phi_X\rangle|\Phi_I\rangle &= \frac{1}{2}\left(|0\rangle|1\rangle|0\rangle|0\rangle + |0\rangle|1\rangle|1\rangle|1\rangle + |1\rangle|0\rangle|0\rangle|0\rangle + |1\rangle|0\rangle|1\rangle|1\rangle\right), \\ &\xrightarrow{CNOT\otimes CNOT} \frac{1}{2}\left(|0\rangle|1\rangle|0\rangle|1\rangle + |0\rangle|1\rangle|1\rangle|0\rangle + |1\rangle|0\rangle|1\rangle|0\rangle + |1\rangle|0\rangle|0\rangle|1\rangle\right), \\ &= |\Phi_X\rangle|\Phi_X\rangle. \end{aligned} \tag{12.15}$$

In this case, the state is no longer left unchanged, and the second pair of qubits is transformed into the state $|\Phi_X\rangle$. However, as we saw in the previous case, this means that in step 2, when Alice and Bob measure, they always obtain different results, and hence they always fail. This is good, since the first pair of qubits was in the state $|\Phi_X\rangle$ (the unwanted state, with flips), and so this procedure allows them to discard these qubits.

Finally, the last case is when they share $|\Phi_X\rangle|\Phi_X\rangle$, which occurs with probability $(1-q)^2$. Here, in step 1 we find

$$\begin{aligned}|\Phi_X\rangle|\Phi_X\rangle &= \frac{1}{2}\left(|0\rangle|1\rangle|0\rangle|1\rangle + |0\rangle|1\rangle|1\rangle|0\rangle + |1\rangle|0\rangle|0\rangle|1\rangle + |1\rangle|0\rangle|1\rangle|0\rangle\right), \\ &\xrightarrow{CNOT\otimes CNOT} \frac{1}{2}\left(|0\rangle|1\rangle|0\rangle|0\rangle + |0\rangle|1\rangle|1\rangle|1\rangle + |1\rangle|0\rangle|1\rangle|1\rangle + |1\rangle|0\rangle|0\rangle|0\rangle\right), \\ &= |\Phi_X\rangle|\Phi_I\rangle. \end{aligned} \tag{12.16}$$

This again changes the state, so that the second pair of qubits end up in the state $|\Phi_I\rangle$. This means that in step 2, when measured, Alice and Bob will always obtain the same outcomes, and hence they will never fail in step 3.

Putting everything together, this means that in the second and third cases Alice and Bob always fail, while in the first and last case, they never fail. The total probability to not fail is therefore $q^2 + (1-q)^2$, and the so the state that they share, in the case of success, will be the ***renormalised*** state

$$\rho' = \frac{q^2|\Phi_I\rangle\langle\Phi_I| + (1-q)^2|\Phi_X\rangle\langle\Phi_X|}{q^2 + (1-q)^2}, \tag{12.17}$$

which is indeed a normalised density operator. We see that the net effect of the protocol was to '***remove***' certain contributions to the density operator, and it can be viewed as a type of probabilistic ***filtering*** process. Notably, the state (12.17) is of the form (12.12), but with a parameter q' equal to

$$q' = \frac{q^2}{q^2 + (1-q)^2}. \tag{12.18}$$

We will have ***improved*** the entanglement in the state if $q' > q$, i.e. if we have decreased the relatively likelihood of having the erroneous state $|\Phi_X\rangle$. A direct calculation (left as an exercise)

shows that $q' > q$ whenever $1/2 < q < 1$, that is ***whenever it is more likely that the state was*** $|\Phi_I\rangle$ ***than*** $|\Phi_X\rangle$***!*** The largest improvement in the quality of the entanglement occurs when $q = 3/4$, in which case $q' = 9/10$, which is 20% bigger (i.e. $q'/q = 1.2$).

Finally, there is nothing to stop us applying this procedure multiple times. That is, we could imagine distributing four pairs of qubits, and applying the above procedure twice, on a pair at a time. If both procedures succeed, we could then apply the procedure a third time, now on the pair of remaining qubits from the two original pairs. If this procedure also succeeds, then we have purified the entanglement twice. Significant gains can be achieved this way, albeit at the price of lowering the overall success probability of the procedure (this is left as an exercise).

The main lesson here is that it is in general possible to probabilistically improve the quality of entanglement in a state, making it much more useful for applications, such as the quantum internet. While we have only studied a specific example, the general lesson indeed holds, and it is generally possible to find (much more complicated) ***purification procedures***, with better success probabilities, and which improve the quality of the entanglement further.