Contents

Introduction 4
Plan 4
Topics 4

Some mathematical preliminaries 6
0.1 Euclidean vectors 6
0.2 Vector spaces 6
0.3 Bras and kets 7
0.4 Daggers 8
0.5 Geometry 8
0.6 Operators 9
0.7 Outer products 10
0.8 The trace 11
0.9 Some useful identities 11

I Foundations 13

1 Quantum interference: an overview 14
1.1 Two basic rules 14
1.2 Quantum interference: the failure of probability theory 15
1.3 Superpositions 17
1.4 Interferometers 17
1.5 Qubits, gates, and circuits 20
1.6 Quantum decoherence 21
1.7 Computation: deterministic, probabilistic, and quantum 23
1.8 Computational complexity 24
1.9 Outlook 26
1.10 Remarks and exercises 26

2 Qubits 33
2.1 Composing quantum operations 33
2.2 Quantum bits, called “qubits” 34
2.3 Quantum gates and circuits 36
2.4 Single qubit interference 36
2.5 The square root of NOT 38
2.6 Phase gates galore 39
2.7 Pauli operators 39
2.8 From bit-flips to phase-flips, and back again 40
2.9 Any unitary operation on a single qubit 40
2.10 The Bloch sphere 41
2.11 Composition of rotations 44
2.12 A finite set of universal gates 44
2.13 Remarks and Exercises 44

3 Logic and geometry with quantum gates **46**
3.1 Physics against logic, via beamsplitters 46
3.2 Quantum interference revisited (still about beam-splitters) 49
3.3 The Pauli matrices, algebraically 51
3.4 Unitaries as rotations 53
3.5 Universality, again 56
3.6 Some quantum dynamics 57
3.7 Remarks and Exercises 58

4 Measurements **61**
4.1 Hilbert spaces, briefly 61
4.2 Back to qubits; complete measurements 61
4.3 The projection rule; incomplete measurements 63
4.4 Example of an incomplete measurement 64
4.5 Observables 65
4.6 Compatible observables and the uncertainty relation 66
4.7 Quantum communication 66
4.8 Basic quantum coding and decoding 67
4.9 Distinguishability of non-orthogonal states 68
4.10 Wiesner's quantum money 70
4.11 Remarks and Exercises 70
4.12 Quantum theory, formally 71

5 Quantum entanglement **74**
5.1 A small history 74
5.2 One, two, many... 74
5.3 Quantum theory, formally (continued) 75
5.4 Back to qubits 77
5.5 Separable or entangled? 77
5.6 Controlled-NOT 78
5.7 Other controlled gates 83
5.8 Why qubits, subsystems, and entanglement? 87
5.9 Remarks and exercises 88
5.10 Appendix: Tensor products in components 91
5.11 Appendix: The Schmidt decomposition 94

6 Density matrices **96**
6.1 Definitions 96
6.2 Mixtures 97
6.3 A few instructive examples, and some less instructive remarks 98
6.4 Mixed states of a qubit, and the Bloch ball 100
6.5 Subsystems of entangled systems 100
6.6 Partial trace, revisited 102
6.7 Mixtures and subsystems 102
6.8 Partial trace, yet again 104
6.9 Remarks and exercises 104

II The power of interference and entanglement **106**

7 Bell's theorem **107**
7.1 Quantum correlations 107
7.2 Hidden variables 108
7.3 CHSH inequality 108
7.4 Quantum correlations revisited 109
7.5 Tsirelson's inequality 110
7.6 Remarks and Exercises 110

8 Quantum algorithms **111**
8.1 Quantum Boolean function evaluation 111
8.2 More phase kick-back 112
8.3 Oracles and query complexity 113
8.4 Three more quantum algorithms 115
8.5 Remarks and exercises 122

9 Decoherence, and elements of quantum error correction **123**
9.1 Decoherence simplified 123
9.2 Decoherence and interference 123
9.3 Evolution of density operators under decoherence 124
9.4 Quantum errors 125
9.5 Same evolution, different errors 126
9.6 Some errors can be corrected on some states 127
9.7 Repetition codes 128
9.8 Quantum error correction 128
9.9 Turning bit-flips into phase-flips 129
9.10 Dealing with bit-flip and phase-flip errors 129
9.11 Remarks and Exercises 129

III Quantum security **130**

10 Quantum channels, or CP maps **131**
10.1 The constructive approach 132
10.2 The axiomatic approach 140
10.3 Comparing the two approaches 145
10.4 What are positive maps good for? 147
10.5 Remarks and exercises 148

11 Quantum error correction and fault tolerance **156**

Introduction

For the past however-many years, Artur Ekert has been teaching the masters course "Introduction to Quantum Information" at the University of Oxford. During this time, many versions of accompanying lecture notes have come and gone, with constant improvements and changes being made. The version that you will find on this website has been carefully edited by Tim Hosgood into a cohesive "book", containing additional exercises and topics. Thanks go to Zhenyu Cai for many helpful comments and corrections, and we also appreciate the work of Yihui Xie in developing the Bookdown package with which this document was built.

For more information, see the accompanying website.

Plan

In this series of lectures you will learn how inherently quantum phenomena, such as quantum interference and quantum entanglement, can make information processing more efficient and more secure, even in the presence of noise.

The interdisciplinary nature of this topic, combined with the diverse backgrounds that different readers have, means that some may find some particular chapters easy, while others find them difficult. The following will be assumed as prerequisites: elementary probability theory, complex numbers, vectors and matrices, tensor products, and Dirac braket notation. A basic knowledge of quantum mechanics (especially in the simple context of finite dimensional state spaces, e.g. state vectors, composite systems, unitary matrices, Born rule for quantum measurements) and some ideas from classical theoretical computer science (complexity theory) would be helpful, but is not at all essential. Some of these things are covered at the end of this chapter.

Topics

- Fundamentals of quantum theory
  - addition of probability amplitudes
  - quantum interference
  - mathematical description of states and evolution of closed quantum systems (Hilbert space, unitary evolution)
  - measurements (projectors, Born rule)
  - Pauli matrices
- Distinguishability of quantum states
- The Bloch sphere
  - parametrisation
  - action of quantum gates on the Bloch vector
- The definition of quantum entanglement (the tensor product structure)
- The no-cloning theorem, and quantum teleportation
- Quantum gates
  - phase gate
  - Hadamard
  - controlled-`NOT`
  - SWAP
  - the Hadamard-phase-Hadamard network

  - phase “kick-back” induced by controlled-U
  - phase “kick-back” induced by quantum Boolean function evaluation
- Quantum algorithms
  - Deutsch
  - Bernstein-Vazirani
  - Simon
- Bell’s theorem
  - Quantum correlations
  - CHSH inequality
- Density matrices
  - partial trace
  - statistical mixture of pure states
  - Born rule for density matrices
  - quantum entanglement in terms of density matrices
- Completely positive maps
  - Kraus operators
  - the Choi matrix
  - positive versus completely positive maps
  - partial-transpose
- The simple model of decoherence
- Quantum error correction of bit-flip and phase-flip errors

Some mathematical preliminaries

0.1 Euclidean vectors

We assume that you are familiar with Euclidean vectors — those arrow-like geometric objects which are used to represent physical quantities, such as velocities, or forces. You know that any two velocities can be added to yield a third, and the multiplication of a "velocity vector" by a real number is another "velocity vector". So a **linear combination** of vectors is another vector. Mathematicians have simply taken these properties and defined vectors as *anything* that we can add and multiply by numbers, as long as everything behaves in a nice enough way. This is basically what an Italian mathematician Giuseppe Peano (1858–1932) did in a chapter of his 1888 book with an impressive title: *Calcolo geometrico secondo l'Ausdehnungslehre di H. Grassmann preceduto dalle operazioni della logica deduttiva*.

0.2 Vector spaces

Following Peano, we define a **vector space** as a mathematical structure in which the notion of linear combination "makes sense".

More formally, a **complex vector space** is a set V such that, given any two **vectors** a and b (that is, any two elements of V) and any two complex numbers α and β, we can form the linear combination $\alpha a + \beta b$, which is also a vector in V.

As we said, there are certain "nice properties" that these things must satisfy. Addition of vectors must be commutative and associative, with an identity (the zero vector, which will always be written as $\mathbf{0}$) and an inverse for each v (written as $-v$). Multiplication by complex numbers must obey the two distributive laws: $(\alpha + \beta)v = \alpha v + \beta v$ and $\alpha(v + w) = \alpha v + \alpha w$.

A **subspace** of V is any subset of V which is closed under vector addition and multiplication by complex numbers. Here we start using the Dirac bra-ket notation and write vectors in a somewhat fancy way as $|\text{label}\rangle$, where the "label" is anything that serves to specify what the vector is. For example, $|\uparrow\rangle$ and $|\downarrow\rangle$ may refer to an electron with spin up or down along some prescribed direction and $|0\rangle$ and $|1\rangle$ may describe a quantum bit (a "qubit") holding either logical 0 or 1. These are often called **ket** vectors, or simply **kets**. (We will deal with "bras" in a moment). A **basis** in V is a collection of vectors $|e_1\rangle, |e_2\rangle, \ldots, |e_n\rangle$ such that every vector $|v\rangle$ in V can be written (in *exactly* one way) as a linear combination of the basis vectors; $|v\rangle = \sum_i v_i |e_i\rangle$. The number of elements in a basis is called the **dimension** of V. (Showing that this definition is independent of the basis that we choose is a "fun" linear algebra exercise). The most common n-dimensional complex vector space is the space of ordered n-tuples of complex numbers, usually written as column vectors:

$$|a\rangle = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} \qquad |b\rangle = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$$

$$\alpha|a\rangle + \beta|b\rangle = \begin{bmatrix} \alpha a_1 + \beta b_1 \\ \alpha a_2 + \beta b_2 \\ \vdots \\ \alpha a_n + \beta b_n \end{bmatrix}$$

In fact, this is the space we will use most of the time. Throughout the course we will deal only with vector spaces of *finite* dimensions. This is sufficient for all our purposes and we will avoid many mathematical subtleties associated with infinite dimensional spaces, for which we would need to tools of **functional analysis**.

0.3 Bras and kets

An **inner product** on a vector space V (over the complex numbers) is a function that assigns to each pair of vectors $|u\rangle, |v\rangle \in V$ a complex number $\langle u|v\rangle$, and satisfies the following conditions:

- $\langle u|v\rangle = \langle v|u\rangle^\star$;
- $\langle v|v\rangle \geqslant 0$ for all $|v\rangle$;
- $\langle v|v\rangle = 0$ if and only if $|v\rangle = 0$.

The inner product must also be *linear* in the second argument but *antilinear* in the first argument:

$$\langle c_1 u_1 + c_2 u_2|v\rangle = c_1^\star \langle u_1|v\rangle + c_2^\star \langle u_2|v\rangle$$

for any complex constants c_1 and c_2.

With any physical system we associate a complex vector space with an inner product, known as a **Hilbert space** $\mathcal{H}$. The inner product between vectors $|u\rangle$ and $|v\rangle$ in $\mathcal{H}$ is written as

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

The term "Hilbert space" used to be reserved for an infinite-dimensional inner product space that is **complete**, i.e. such that every Cauchy sequence in the space converges to an element in the space. Nowadays, as in these notes, the term includes finite-dimensional spaces, which automatically satisfy the condition of completeness.

For example, for column vectors $|u\rangle$ and $|v\rangle$ in $\mathbb{C}^n$ written as

$$|u\rangle = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} \qquad |v\rangle = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$$

their inner product is defined as

$$\langle u|v\rangle = u_1^\star v_1 + u_2^\star v_2 + \ldots + u_n^\star v_n.$$

Following Dirac we may split the inner product into two ingredients

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

Here $|v\rangle$ is a ket vector, and $\langle u|$ is called a **bra** vector, or a **bra**, and can be represented by a row vector:

$$\langle u| = \begin{bmatrix} u_1^\star, & u_2^\star, & \ldots, & u_n^\star \end{bmatrix}.$$

The inner product can now be viewed as the result of the matrix multiplication:

$$\begin{aligned} \langle u|v\rangle &= \begin{bmatrix} u_1^\star, & u_2^\star, & \ldots, & u_n^\star \end{bmatrix} \cdot \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} \\ &= u_1^\star v_1 + u_2^\star v_2 + \ldots + u_n^\star v_n. \end{aligned}$$

Bras are vectors: you can add them, and multiply them by scalars (which, here, are complex numbers), but they are vectors in the space $\mathcal{H}^\star$ which is **dual** to $\mathcal{H}$. Elements of

$\mathcal{H}^\star$ are **linear functionals**, that is, linear maps from $\mathcal{H}$ to $\mathbb{C}$. A linear functional $\langle u|$ acting on a vector $|v\rangle$ in $\mathcal{H}$ gives a complex number $\langle u|v\rangle$.

All Hilbert spaces of the same dimension are isomorphic, so the differences between quantum systems cannot be really understood without additional structure. This structure is provided by a specific algebra of operators acting on $\mathcal{H}$.

0.4 Daggers

Although $\mathcal{H}$ and $\mathcal{H}^\star$ are not identical spaces – the former is inhabited by kets, and the latter by bras – they are closely related. There is a bijective map from one to the other, $|v\rangle \leftrightarrow \langle v|$, denoted by a **dagger**:

"Is this a † which I see before me…"

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

We usually omit the parentheses when it is obvious what the dagger operation applies to.

The dagger operation, also known as **Hermitian conjugation**, is *antilinear*:

$$(c_1|v_1\rangle + c_2|v_2\rangle)^\dagger = c_1^\star\langle v_1| + c_2^\star\langle v_2|$$
$$(c_1\langle v_1| + c_2\langle v_2|)^\dagger = c_1^\star|v_1\rangle + c_2^\star|v_2\rangle.$$

Also, when applied twice, the dagger operation is the identity map. In the matrix representation,

$$|v\rangle = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} \overset{\dagger}{\longleftrightarrow} \langle v| = \begin{bmatrix} v_1^\star, & v_2^\star, & \ldots, & v_n^\star \end{bmatrix}.$$

Recall that the conjugate transpose, or the Hermitian conjugate, of an $(n \times m)$ matrix A is an $(m \times n)$ matrix $A^\dagger$, obtained by interchanging the rows and columns of A and taking complex conjugates of each entry in A, i.e. $A^\dagger_{ij} = A^\star_{ji}$. In mathematics texts it is often denoted by * rather than †.

0.5 Geometry

The inner product brings geometry: the **length**, or **norm**, of $|v\rangle$ is given by $\|v\| = \sqrt{\langle v|v\rangle}$, and we say that $|u\rangle$ and $|v\rangle$ are **orthogonal** if $\langle u|v\rangle = 0$. Any maximal set of pairwise orthogonal vectors of unit length forms an orthonormal basis, and so any vector can be expressed as a linear combination of the basis vectors:

That is, consider sets of vectors $|e_i\rangle$ such that $\langle e_i|e_j\rangle = \delta_{ij}$ (where the **Kronecker delta** δ_{ij} is 0 if $i \neq j$, and 1 if $i = j$.), and then pick any of the largest such sets (which must exist, since we assume our vector spaces to be finite dimensional).

$$|v\rangle = \sum_i v_i|e_i\rangle$$
where $v_i = \langle e_i|v\rangle$.

Then the bras $\langle e_i|$ form the **dual basis**

$$\langle v| = \sum_i v_i^\star\langle e_i|$$
where $v_i^\star = \langle v|e_i\rangle$.

To make the notation a bit less cumbersome, we will sometimes label the basis kets as

$|i\rangle$ rather than $|e_i\rangle$, and write

$$|v\rangle = \sum_i |i\rangle\langle i|v\rangle$$
$$\langle v| = \sum_j \langle v|i\rangle\langle i|.$$

But *do not confuse* $|0\rangle$ *with the zero vector*! We *never* write the zero vector as $|0\rangle$, but only ever as 0, without any bra or ket decorations (so e.g. $|v\rangle + 0 = |v\rangle$).

With any *isolated* quantum system, which can be prepared in n *perfectly distinguishable* states, we can associate a Hilbert space $\mathcal{H}$ of dimension n such that each vector $|v\rangle \in \mathcal{H}$ of unit length (i.e. $\langle v|v\rangle = 1$) represents a quantum state of the system.
The overall phase of the vector has no physical significance: $|v\rangle$ and $e^{i\varphi}|v\rangle$ (for any real φ) both describe the same state.
The inner product $\langle u|v\rangle$ is the *probability amplitude* that a quantum system prepared in state $|v\rangle$ will be found in state $|u\rangle$ upon measurement.
States corresponding to orthogonal vectors (i.e. $\langle u|v\rangle = 0$) are *perfectly distinguishable*, since, if we prepare the system in state $|v\rangle$, then it will never be found in state $|u\rangle$, and vice versa. In particular, states forming orthonormal bases are always perfectly distinguishable from each other. Choosing such states, as we shall see in a moment, is equivalent to choosing a particular quantum measurement.

0.6 Operators

A **linear map** between two vector spaces $\mathcal{H}$ and $\mathcal{K}$ is a function $A\colon \mathcal{H} \to \mathcal{K}$ that respects linear combinations:

$$A(c_1|v_1\rangle + c_2|v_2\rangle) = c_1 A|v_1\rangle + c_2 A|v_2\rangle$$

for any vectors $|v_1\rangle, |v_2\rangle$ and any complex numbers c_1, c_2. We will focus mostly on **endomorphisms**, that is, maps from $\mathcal{H}$ to $\mathcal{H}$, and we will call them **operators**. The symbol $\mathbf{1}$ is reserved for the identity operator that maps every element of $\mathcal{H}$ to itself (i.e. $\mathbf{1}|v\rangle = |v\rangle$ for all $|v\rangle \in \mathcal{H}$). The product AB of two operators A and B is the operator obtained by first applying B to some ket $|v\rangle$ and then A to the ket which results from applying B:

$$(AB)|v\rangle = A(B|v\rangle).$$

The order *does* matter: in general, $AB \neq BA$. In the exceptional case in which $AB = BA$, one says that these two operators **commute**. The inverse of A, written as A^{-1}, is the operator that satisfies $AA^{-1} = \mathbf{1} = A^{-1}A$. For finite-dimensional spaces, one only needs to check *one* of these two conditions, since any one of the two implies the other, whereas, on an infinite-dimensional space, *both* must be checked. Finally, given a particular basis, an operator A is uniquely determined by the entries of its matrix, defined by $A_{ij} = \langle i|A|j\rangle$. The **adjoint**, or **Hermitian conjugate**, of A, denoted by $A^\dagger$, is defined by the relation

$$\langle i|A^\dagger|j\rangle = \langle j|A|i\rangle^\star$$
for all $|i\rangle, |j\rangle \in \mathcal{H}$.

An operator A is said to be

- **normal** if $AA^\dagger = A^\dagger A$,
- **unitary** if $AA^\dagger = A^\dagger A = \mathbf{1}$,
- **Hermitian** (or **self-adjoint**) if $A^\dagger = A$.

Any physically admissible evolution of an isolated quantum system is represented by a unitary operator. Note that unitary operators preserve the inner product: given a unitary operator U and two kets $|a\rangle$ and $|b\rangle$, and defining $|a'\rangle = U|a\rangle$ and $|b'\rangle = U|b\rangle$, we have that

$$\begin{aligned} \langle a'| &= \langle a|U^\dagger \\ \langle b'| &= \langle b|U^\dagger \\ \langle a'|b'\rangle = \langle a|U^\dagger U|b\rangle &= \langle a|\mathbf{1}|b\rangle = \langle a|b\rangle. \end{aligned}$$

Preserving the inner product implies preserving the norm induced by this product, i.e. unit state vectors are mapped to unit state vectors, i.e. *unitary operations are the isometries of the Euclidean norm*.

0.7 Outer products

Apart from the inner product $\langle u|v\rangle$, which is a complex number, we can also form the **outer product** $|u\rangle\langle v|$, which is a linear map (operator) on $\mathcal{H}$ (or on $\mathcal{H}^\star$, depending how you look at it). This is what physicists like (and what mathematicians dislike!) about Dirac notation: a certain degree of healthy ambiguity.

- The result of $|u\rangle\langle v|$ acting on a ket $|x\rangle$ is $|u\rangle\langle v|x\rangle$, i.e. the vector $|u\rangle$ multiplied by the complex number $\langle v|x\rangle$.
- Similarly, the result of $|u\rangle\langle v|$ acting on a bra $\langle y|$ is $\langle y|u\rangle\langle v|$, i.e. the functional $\langle v|$ multiplied by the complex number $\langle y|u\rangle$.

The product of two maps, $A = |a\rangle\langle b|$ followed by $B = |c\rangle\langle d|$, is a linear map BA, which can be written in Dirac notation as

$$BA = |c\rangle\langle d|a\rangle\langle b| = \langle d|a\rangle|c\rangle\langle b|$$

i.e. the inner product (complex number) $\langle d|a\rangle$ times the outer product (linear map) $|c\rangle\langle b|$.

Any operator on $\mathcal{H}$ can be expressed as a sum of outer products. Given an orthonormal basis $\{|e_i\rangle\}$, any operator which maps the basis vectors $|e_i\rangle$ to vectors $|f_i\rangle$ can be written as $\sum_i |f_i\rangle\langle e_i|$, where the sum is over all the vectors in the orthonormal basis. If the vectors $\{|f_i\rangle\}$ also form an orthonormal basis then the operator simply "rotates" one orthonormal basis into another. These are unitary operators which preserve the inner product. In particular, if each $|e_i\rangle$ is mapped to $|e_i\rangle$, then we obtain the identity operator:

$$\sum_i |e_i\rangle\langle e_i| = \mathbf{1}.$$

This relation holds for *any* orthonormal basis, and it is one of the most ubiquitous and useful formulas in quantum theory. For example, for any vector $|v\rangle$ and for any orthonormal

basis $\{|e_i\rangle\}$, we have

$$\begin{aligned}|v\rangle &= \mathbf{1}|v\rangle \\ &= \sum_i |e_i\rangle\langle e_i|\ |v\rangle \\ &= \sum_i |e_i\rangle\ \langle e_i|v\rangle \\ &= \sum_i v_i|e_i\rangle,\end{aligned}$$

where $v_i = \langle e_i|v\rangle$ are the components of $|v\rangle$. Finally, note that the adjoint of $|a\rangle\langle b|$ is $|b\rangle\langle a|$.

0.8 The trace

The **trace** is an operation which turns outer products into inner products,

$$\mathrm{tr}\colon |b\rangle\langle a| \longmapsto \langle a|b\rangle.$$

We have just seen that any linear operator can be written as a sum of outer products, and so we can extend the definition of trace (by linearity) to any operator. Alternatively, for any square matrix A, the trace of A is defined to be the sum of its diagonal elements:

$$\mathrm{tr}\, A = \sum_k \langle e_k|A|e_k\rangle = \sum_k A_{kk}.$$

You can show, using this definition or otherwise, that the trace is cyclic (i.e. $\mathrm{tr}(AB) = \mathrm{tr}(BA)$) and linear (i.e. $\mathrm{tr}(\alpha A + \beta B) = \alpha\,\mathrm{tr}(A) + \beta\,\mathrm{tr}(B)$, where A and B are square matrices and α and β complex numbers). Moreover,

$$\begin{aligned}\mathrm{tr}\,|b\rangle\langle a| &= \sum_k \langle e_k|b\rangle\langle a|e_k\rangle \\ &= \sum_k \langle a|e_k\rangle\langle e_k|b\rangle \\ &= \langle a|\mathbf{1}\rangle|b\rangle \\ &= \langle a|b\rangle.\end{aligned}$$

Here, the second term can be viewed both as the sum of the diagonal elements of $|b\rangle\langle a|$ in the $|e_k\rangle$ basis, and as the sum of the products of two complex numbers $\langle e_k|b\rangle$ and $\langle a|e_k\rangle$. We have used the decomposition of the identity, $\sum_k |e_k\rangle\langle e_k| = \mathbf{1}$. Given that we can decompose the identity by choosing any orthonormal basis, it is clear that the trace does *not* depend on the choice of the basis.

0.9 Some useful identities

- $|a\rangle^\dagger = \langle a|$
- $\langle a|^\dagger = |a\rangle$
- $(\alpha|a\rangle + \beta|b\rangle)^\dagger = \alpha^\star\langle a| + \beta^\star\langle b|$
- $(|a\rangle\langle b|)^\dagger = |b\rangle\langle a|$
- $(AB)^\dagger = B^\dagger A^\dagger$
- $(\alpha A + \beta B)^\dagger = \alpha^\star A^\dagger + \beta^\star B^\dagger$
- $(A^\dagger)^\dagger = A$
- $\mathrm{tr}(\alpha A + \beta B) = \alpha\,\mathrm{tr}(A) + \beta\,\mathrm{tr}(B)$

- $\mathrm{tr}\,|a\rangle\langle b| = \langle b|a\rangle$
- $\mathrm{tr}(ABC) = \mathrm{tr}(CAB) = \mathrm{tr}(BCA)$

Foundations

PART

I

1 Quantum interference: an overview

*About complex numbers, called **probability amplitudes**, that, unlike probabilities, can cancel each other out, leading to **quantum interference**, and consequently qualitatively new ways of processing information.*

The classical theory of computation does not usually refer to physics. Pioneers such as Alan Turing, Alonzo Church, Emil Post, and Kurt Gödel managed to capture the correct classical theory by intuition alone and, as a result, it is often falsely assumed that its foundations are self-evident and purely abstract. They are not!

Computation is a physical process.
Computation is a physical process.
Computation is . . .

The concepts of information and computation can be properly formulated only in the context of a physical theory — information is stored, transmitted and processed always by *physical* means. Computers are physical objects and computation is a physical process. Indeed, any computation, classical or quantum, can be viewed in terms of physical experiments, which produce **outputs** that depend on initial preparations called **inputs**. Once we abandon the classical view of computation as a purely logical notion independent of the laws of physics it becomes clear that whenever we improve our knowledge about physical reality, we may also gain new means of computation. Thus, from this perspective, it is not very surprising that the discovery of quantum mechanics in particular has changed our understanding of the nature of computation. In order to explain what makes quantum computers so different from their classical counterparts, we begin with the rudiments of quantum theory.

Some of what we say in this chapter will be repeated in later chapters, but usually in much more detail. Feel free to think of this chapter as a sort of "aeroplane tour" of the rudiments, knowing that we will soon land on the ground to go out exploring by foot.

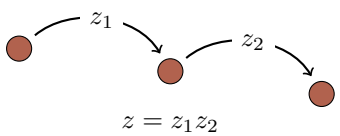
1.1 Two basic rules

Quantum theory, at least at some instrumental level, can be viewed as a modification of probability theory. We replace positive numbers (probabilities) with complex numbers z (called **probability amplitudes**) such that the squares of their absolute values, $|z|^2$, are interpreted as probabilities.

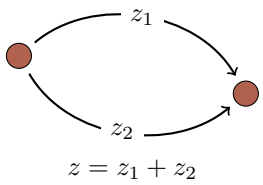
> The correspondence between probability amplitudes z and probabilities $|z|^2$ is known as **Born's Rule**.

The rules for combining amplitudes are very reminiscent of the rules for combining probabilities:

1. Whenever something can happen in a sequence of independent steps, we multiply the amplitudes of each step.



2. Whenever something can happen in several alternative ways, we add the amplitudes for each separate way.



That's it! These two rules are basically all you need to manipulate amplitudes in any physical process, no matter how complicated. They are universal and apply to any physical system, from elementary particles through atoms and molecules to white dwarfs stars. They also apply to information, since, as we have already emphasised, information is physical. The two rules look deceptively simple but, as you will see in a moment, their consequences are anything but trivial.

We will, however, amend the two rules later on when we touch upon particle statistics.

1.2 Quantum interference: the failure of probability theory

Modern mathematical probability theory is based on three axioms, proposed by Andrey Nikolaevich Kolmogorov (1903–1987) in his monograph with the impressive German title *Grundbegriffe der Wahrscheinlichkeitsrechnung* ("Foundations of Probability Theory"). The **Kolmogorov axioms** are simple and intuitive:

I always found it an interesting coincidence that the two basic ingredients of modern quantum theory, namely probability and complex numbers, were discovered by the same person, an extraordinary man of many talents: a gambling scholar by the name of Girolamo Cardano (1501–1576).

1. Once you identify all elementary outcomes, or events, you may then assign probabilities to them.
2. Probability is a number between 0 and 1, and an event which is certain has probability 1.
3. Last but not least, the probability of any event can be calculated using a deceptively simple rule — the **additivity axiom**: *Whenever an event can occur in several mutually exclusive ways, the probability for the event is the sum of the probabilities for each way considered separately.*

Obvious, isn't it? So obvious, in fact, that probability theory was accepted as a mathematical framework theory, a language that can be used to describe actual physical phenomena. Physics should be able to identify elementary events and assign numerical probabilities to them. Once this is done we may revert to mathematical formalism of probability theory. The Kolmogorov axioms will take care of the mathematical consistency and will guide us whenever there is a need to calculate probabilities of more complex events. This is a very sensible approach, apart from the fact that it does not always work! Today, we know that probability theory, as ubiquitous as it is, fails to describe many common quantum phenomena. In order to see the need for quantum theory let us consider a simple experiment in which probability theory fails to give the right predictions.

1.2.1 The double slit experiment

In a double slit experiment, a particle emitted from a source S can reach the detector D by taking two different paths, e.g. through an upper or a lower slit in a barrier between the source and the detector. After sufficiently many repetitions of this experiment we can evaluate the frequency of clicks in the detector D and show that it is inconsistent with the predictions based on probability theory. Let us use the quantum approach to show how the discrepancy arises.

The particle emitted from a source S can reach detector D by taking two different paths, with amplitudes z_1 and z_2 respectively. We may say that the upper slit is taken with probability $p_1 = |z_1|^2$ and the lower slit with probability $p_2 = |z_2|^2$. These are two mutually

exclusive events. With the two slits open, probability theory declares (by the additivity axiom) that the particle should reach the detector with probability $p_1 + p_2 = |z_1|^2 + |z_2|^2$. But this is not what happens experimentally!

Following the "quantum rules", first we add the amplitudes and then we square the absolute value of the sum to get the probability. Thus, the particle will reach the detector with probability

$$\begin{aligned} p &= |z|^2 \\ &= |z_1 + z_2|^2 \\ &= |z_1|^2 + |z_2|^2 + z_1^\star z_2 + z_1 z_2^\star \\ &= p_1 + p_2 + |z_1||z_2| \left(e^{i(\varphi_2 - \varphi_1)} + e^{-i(\varphi_2 - \varphi_1)} \right) \\ &= p_1 + p_2 + 2\sqrt{p_1 p_2}\cos(\varphi_2 - \varphi_1) \\ &= p_1 + p_2 + \text{interference terms} \end{aligned} \tag{1.2.1.1}$$

where we have expressed the amplitudes in their polar forms

$$\begin{aligned} z_1 &= |z_1| e^{i\varphi_1} \\ z_2 &= |z_2| e^{i\varphi_2}. \end{aligned}$$

The appearance of the interference terms marks the departure from the classical theory of probability. The probability of any two seemingly mutually exclusive events is the sum of the probabilities of the individual events, $p_1 + p_2$, *modified* by the **interference term** $2\sqrt{p_1 p_2}\cos(\varphi_2 - \varphi_1)$. Depending on the **relative phase** $\varphi_2 - \varphi_1$, the interference term can be either negative (which we call **destructive** interference) or positive (**constructive** interference), leading to either suppression or enhancement of the total probability p.

The algebra is simple; our focus is on the physical interpretation. Firstly, note that the important quantity here is the *relative* phase $\varphi_2 - \varphi_1$ rather than the individual values φ_1 and φ_2. This observation is not trivial at all: if a particle reacts only to the difference of the two phases, each pertaining to a separate path, then it must have, somehow, experienced the two paths, right? Thus we cannot say that the particle has travelled *either* through the upper or the lower slit, because it has travelled through *both*. In the same way, quantum computers follow, in some tangible way, *all* computational paths simultaneously, producing answers that depend on *all* these alternative calculations. Weird, but this is how it is!

Secondly, what has happened to the additivity axiom in probability theory? What was wrong with it? One problem is the assumption that the processes of taking the upper or the lower slit are mutually exclusive; in reality, as we have just mentioned, the two transitions *both occur*, simultaneously. However, we cannot learn this from probability theory, nor from any other *a priori* mathematical construct.

There is no fundamental reason why Nature should conform to the additivity axiom.

We find out how nature works by making intelligent guesses, running experiments, checking what happens and formulating physical theories. If our guess disagrees with experiments then it is wrong, so we try another intelligent guess, and another, etc. Right now, quantum theory is the best guess we have: it offers good explanations and predictions that have not been falsified by any of the existing experiments. This said, rest assured that one day quantum theory *will* be falsified, and then we will have to start guessing all over

According to the philosopher Karl Popper (1902–1994) a theory is genuinely scientific only if it is possible, in principle, to establish that it is false. Genuinely scientific theories are never finally confirmed because no matter how many confirming observations have been made observations that are inconsistent with the empirical predictions of the theory are always possible.

again.

1.3 Superpositions

Amplitudes are more than just tools for calculating probabilities: they tell us something about physical reality. When we deal with probabilities, we may think about them as numbers that quantify our lack of knowledge. Indeed, when we say that a particle goes through the upper or the lower slit with some respective probabilities, it does go through one of the two slits, we just do not know which one. In contrast, according to quantum theory, a particle that goes through the upper and the lower slit with certain amplitudes does explore *both* of the two paths, not just one of them. This is a statement about a real physical situation — about something that is out there and something we can experiment with.

> The assumption that the particle goes through one of the two slits, but just that we do not know which one, is inconsistent with *many* experimental observations.

We have to accept that, apart from some easy to visualise states, known as the **basis states**, (such as the particle at the upper slit or the particle at the lower slit), there are infinitely many other states, all of them equally real, in which the particle is in a **superposition** of the two basis states. This rather bizarre picture of reality is the best we have at the moment, and it works, at least for now.

Physicists write such superposition states as

$$|\psi\rangle = \alpha|\text{at the upper slit}\rangle + \beta|\text{at the lower slit}\rangle,$$

meaning the particle at the upper slit with amplitude α and at the lower slit with amplitude β. Mathematically, you can think about this expression as a vector $|\psi\rangle$ in a two-dimensional complex vector space written in terms of the two basis vectors $|\text{at the upper slit}\rangle$ and $|\text{at the lower slit}\rangle$. You could also write this vector as a column vector with two complex entries α and β, but then you would have to explain the *physical meaning* of the basis states. Here, we use the $|\cdot\rangle$ notation, introduced by Paul Dirac in the early days of the quantum theory as a useful way to write and manipulate vectors. In Dirac notation you can put into the box $|\ \rangle$ anything that serves to specify what the vector is: it could be $|\uparrow\rangle$ for spin up and $|\downarrow\rangle$ for spin down, or $|0\rangle$ for a quantum bit holding logical 0 and $|1\rangle$ for a quantum bit holding logical 1, etc. As we shall see soon, there is much more to this notation, and learning to manipulate it will help you greatly.

Dirac notation will likely be familiar to physicists, but may look odd to mathematicians or computer scientists. Love it or hate it (and I suggest the former), the notation is so common that you simply have no choice but to learn it, especially if you want to study anything related to quantum theory.

1.4 Interferometers

Many modern interference experiments are performed using internal degrees of freedom of atoms and ions. For example, **Ramsey interferometry**, named after American physicist Norman Ramsey, is a generic name for an interference experiment in which atoms are sent through two separate resonant interaction zones, known as **Ramsey zones**, separated by an intermediate dispersive interaction zone.

Many beautiful experiments of this type were carried out in the 1990s in Serge Haroche's lab at the Ecole Normale Supérieure in Paris. Rubidium atoms were sent through two separate interaction zones (resonant interaction in the first and the third cavity) separated by a phase inducing dispersive interaction zone (the central cavity). The atoms were subsequently measured, via a selective ionisation, and found to be in one of the two preselected

energy states, here labeled as $|0\rangle$ and $|1\rangle$. The fraction of atoms found in states $|0\rangle$ or $|1\rangle$ showed a clear dependence on the phase shifts induced by the dispersive interaction in the central cavity. In 2012, Serge Haroche and Dave Wineland shared the Nobel Prize in physics for "ground-breaking experimental methods that enable measuring and manipulation of individual quantum systems."

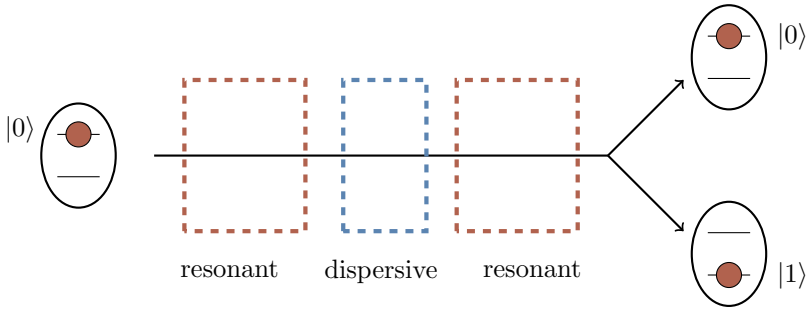


Figure 1. A schematic diagram of a Ramsey interference experiment.

The three rectangular boxes in Figure 1 represent three cavities, each cavity being an arrangement of mirrors which traps electromagnetic field (think about standing waves in between two mirrors). The oval shapes represent rubidium atoms with two preselected energy states labelled as $|0\rangle$ and $|1\rangle$. Each atom is initially prepared in a highly excited internal energy state $|0\rangle$ and zips through the three cavities, from the left to the right. In each cavity the atom interacts with the cavity field. The first and the third cavities are, for all theoretical purposes, identical: their frequencies are tuned to the resonant frequency of the atom, and the atom exchanges energy with the cavity, going back and forth between its energy states $|0\rangle$ and $|1\rangle$. In contrast, in the second (central) cavity, the atom undergoes the so-called dispersive interaction: it is too off-resonance to exchange energy with the field but its energy states "feel" the field and acquire phase shifts. After experiencing this well timed sequence of resonant–dispersive–resonant interactions, the energy of the atom is measured and the atom is found to be either in state $|0\rangle$ or state $|1\rangle$. The fraction of atoms found in state $|0\rangle$ or $|1\rangle$ shows a clear dependence on the phase shifts induced by the dispersive interaction in the central cavity.

We can understand this interference better if we follow the two internal states of the atom as it moves through the three cavities.

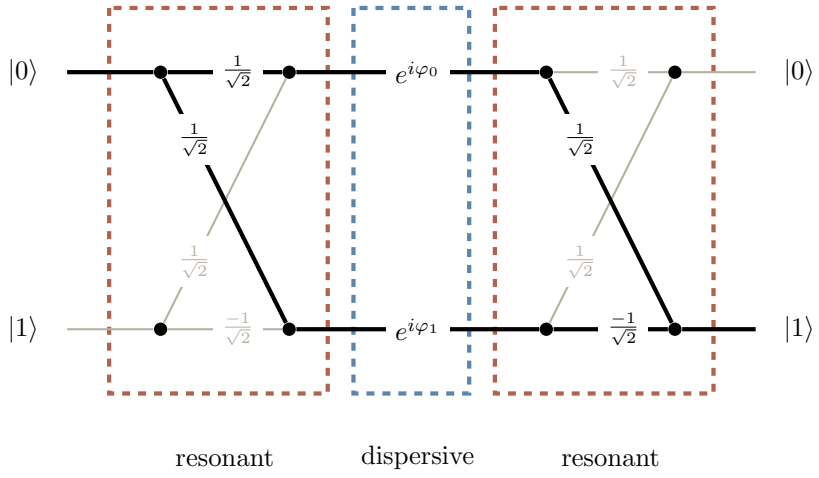


Figure 2. The Ramsey interferometer represented as an abstract diagram. It should be read from left to right. The line segments represent transitions between the two states, $|0\rangle$ and $|1\rangle$, and the numbers are the corresponding probability amplitudes.

Suppose we are interested in the probability that the atom, initially in state $|0\rangle$, will be found, after completing its journey through the three cavities, in state $|1\rangle$. As you can see in Figure 2, this can happen in two ways, as indicated by the two red paths connecting the input state $|0\rangle$ on the left with the output state $|1\rangle$ on the right. Again, let U_{ij} denote the probability amplitude that input $|j\rangle$ generates output $|i\rangle$ (for $i, j = 0, 1$). We can see from the diagram that

$$\begin{aligned} U_{10} &= \frac{1}{\sqrt{2}} e^{i\varphi_0} \frac{1}{\sqrt{2}} + \frac{1}{\sqrt{2}} e^{i\varphi_1} \frac{-1}{\sqrt{2}} \\ &= \frac{1}{2} e^{i\varphi_0} - \frac{1}{2} e^{i\varphi_1} \\ &= -ie^{i\varphi/2} \sin \frac{\varphi}{2}, \end{aligned}$$

where $\varphi = \varphi_0 - \varphi_1$ is the relative phase. The corresponding probability reads

$$\begin{aligned} P_{10} &= |U_{10}|^2 \\ &= \left| \frac{1}{2} e^{i\varphi_0} - \frac{1}{2} e^{i\varphi_1} \right|^2 \\ &= \frac{1}{2} - \frac{1}{2} \cos \varphi. \end{aligned}$$

From the classical probability theory perspective the resonant interaction induces a random switch between $|0\rangle$ and $|1\rangle$ (why?) and the dispersive interaction has no effect on these two states (why?). Hence, one random switch followed by another random switch gives exactly a single random switch, which gives $\frac{1}{2}$ for the probability that input $|0\rangle$ becomes output $|1\rangle$.

You should recognise the first term, $\frac{1}{2}$, as the "classical" probability and the second one, $-\frac{1}{2}\cos\varphi$, as the interference term. We can repeat such calculations for any other pair of input–output states. This approach works fine here but, in general, tracking all possible paths in evolving quantum systems can become messy when the number of input and output states increases. There is, however, a neat way of doing it via matrix multiplication.

The effect of each interaction on atomic states can be described by a matrix of transition amplitudes, as illustrated in Figure 3. Then a sequence of independent interactions is

described by the product of these matrices.

$$U = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} e^{i\varphi_0} & 0 \\ 0 & e^{i\varphi_1} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \end{bmatrix}$$
$$= e^{i\frac{\varphi_0+\varphi_1}{2}} \begin{bmatrix} \cos\frac{\varphi}{2} & -i\sin\frac{\varphi}{2} \\ -i\sin\frac{\varphi}{2} & \cos\frac{\varphi}{2} \end{bmatrix},$$

where $\varphi = \varphi_0 - \varphi_1$.

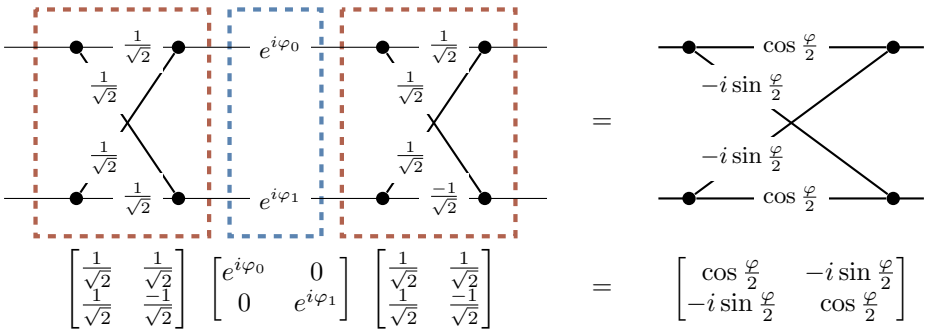
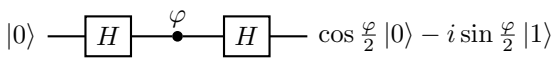


Figure 3. The Ramsey interferometer represented as an abstract diagram (matrix approach). Here we have omitted the $|0\rangle$ and $|1\rangle$ labels, just to simply the diagram. We also ignore the global phase factor of $e^{i\frac{\varphi_0+\varphi_1}{2}}$.

In general, quantum operation A followed by another quantum operation B is a quantum operation described by the matrix product BA (watch the order of matrices). Indeed, the expression $(BA)_{ij} = \sum_k B_{ik}A_{kj}$ is the sum over amplitudes that input $|j\rangle$ generates output $|i\rangle$ via a specific intermediate state $|k\rangle$. As you can see, the matrix approach is a wonderful bookkeeping tool for in one swap it takes care of both multiplying and adding probability amplitudes corresponding to all the contributing paths.

1.5 Qubits, gates, and circuits

Atoms, trapped ions, molecules, nuclear spins and many other quantum objects with two pre-selected basis states labeled as $|0\rangle$ and $|1\rangle$ (from now on we will call such objects quantum bits or **qubits**) can be used to implement simple quantum interference. There is no need to learn about physics behind these diverse technologies if all you want is to understand the basics of quantum theory. We may now conveniently forget about any specific experimental realisation of a qubit and represent a generic **single qubit interference** graphically as a **circuit diagram**:



Do not confuse the interference diagrams of Figure 1 and Figure 3 with the circuit diagram. In the circuit diagrams, which we will use a lot from now on, a single qubit is represented by a single line.

This diagram should be read from left to right. The horizontal line represents a qubit that is inertly carried from one quantum operation to another. We often call this line a **quantum wire**. The wire may describe translation in space (e.g. atoms travelling through cavities) or translation in time (e.g. a sequence of operations performed on a trapped ion). The boxes or circles on the wire represent elementary quantum operations, called **quantum logic gates**. Here we have two types of gates: two Hadamard gates H (think about resonant interactions) and one phase gate P_φ (think about dispersive interaction),

where

$$H = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \end{bmatrix} \quad \text{and} \quad P_\varphi = \begin{bmatrix} 1 & 0 \\ 0 & e^{i\varphi} \end{bmatrix}.$$

Global phase factors are irrelevant, it is the relative phase $\varphi = \varphi_1 - \varphi_0$ that matters. In a single qubit phase gate we usually factor out $e^{i\varphi_0}$, which leaves us with the two diagonal entries: 1 and $e^{i\varphi}$.

The input qubits appear as state vectors on the left side of circuit diagrams, and the output qubits as state vectors on the right. The product of the three matrices $HP_\varphi H$ (see Figure 3) describes the action of the whole circuit: it maps input state vectors to output state vectors:

$$HP_\varphi H = \begin{bmatrix} \cos\frac{\varphi}{2} & -i\sin\frac{\varphi}{2} \\ -i\sin\frac{\varphi}{2} & \cos\frac{\varphi}{2} \end{bmatrix}$$

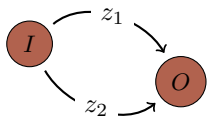
$$\begin{aligned} |0\rangle &\longmapsto \cos\tfrac{\varphi}{2}|0\rangle - i\sin\tfrac{\varphi}{2}|1\rangle, \\ |1\rangle &\longmapsto -i\sin\tfrac{\varphi}{2}|0\rangle + \cos\tfrac{\varphi}{2}|1\rangle. \end{aligned}$$

1.6 Quantum decoherence

We do need quantum theory to describe many physical phenomena, but, at the same time, there are many other phenomena where the classical theory of probability works pretty well. We hardly see quantum interference on a daily basis. Why? The answer is **decoherence**. The addition of probability amplitudes, rather than probabilities, applies to physical systems which are completely isolated. However, it is almost impossible to isolate a complex quantum system, such as a quantum computer, from the rest of the world. There will always be spurious interactions with the environment, and when we add amplitudes, we have to take into account not only different configurations of the physical system at hand, but also different configurations of the environment.

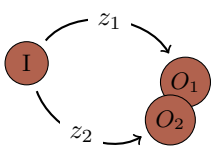
For example, consider an isolated system composed of a quantum computer and its environment. The computer is prepared in some input state I and generates output O. Let us look at the following two scenarios:

1. *The computer is isolated and quantum computation does not affect the environment.* The computer and the environment evolve independently from each other and, as a result, the environment does not hold any physical record of how the computer reached output O. In this case we add the amplitudes for each of the two alternative computational paths.



$$p = |z_1 + z_2|^2$$

2. *Quantum computation affects the environment.* The environment now holds a physical record of how the computer reached output O, which results in two final states of the composed system (computer + environment) which we denote O_1 and O_2. We add the probabilities for each of the two alternative computational paths.

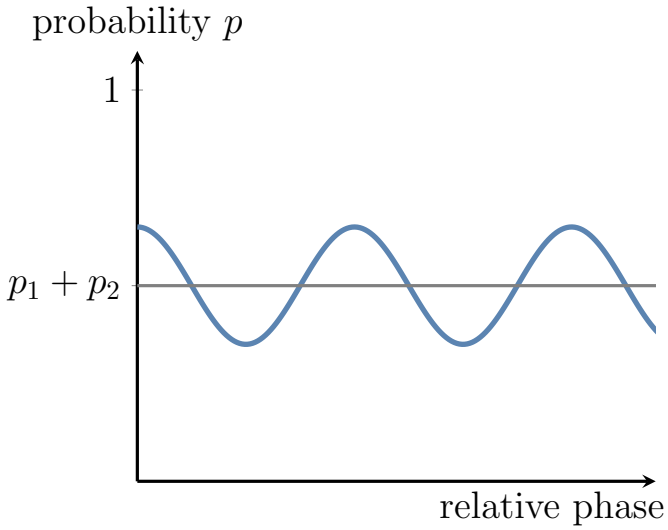


$$p = |z_1|^2 + |z_2|^2$$

When quantum computation affects the environment, we have to include the environment in our analysis for it now takes part in the computation. Depending on which computational path was taken, the environment may end up in two distinct states. The computer itself may show output O, but when we include the environment we have not one, but two outputs, O_1 and O_2, denoting, respectively, "computer shows output O and the environment knows that path 1 was taken" and "computer shows output O and the environment knows that path 2 was taken". There are no alternative ways of reaching O_1 or O_2, hence there is no interference, and the corresponding probabilities read $p_1 = |z_1|^2$ for O_1, and $p_2 = |z_2|^2$ for O_2. The probability that the computer shows output O, regardless the state of the environment, is the sum of of the two probabilities: $p = p_1 + p_2$. We have lost the interference term and any advantages of quantum computation along with it. In the presence of decoherence, the interference formula in Equation (1.2.1.1) is modified and reads

$$p = p_1 + p_2 + 2v\sqrt{p_1 p_2}\cos(\varphi_2 - \varphi_1),$$

where the parameter v, called the **visibility** of the interference pattern, ranges from 0 (the environment can perfectly distinguish between the two paths, total decoherence, no interference) to 1 (the environment cannot distinguish between the two paths, no decoherence, full interference), with the values in between corresponding to partial decoherence.



We shall derive this formula later on, and you will see that v quantifies the degree of distinguishability between O_1 and O_2. The more the environment knows about which path was taken the less interference we see.

Decoherence suppresses quantum interference.

Decoherence is chiefly responsible for our classical description of the world — without interference terms we may as well add probabilities instead of amplitudes. While decoherence is a serious impediment to building quantum computers, depriving us of the power of quantum interference, it is not all doom and gloom: there are clever ways around decoherence, such as quantum error correction and fault-tolerant methods we will meet later.

1.7 Computation: deterministic, probabilistic, and quantum

Take one physical bit or a qubit. It has two logical states: $|0\rangle$ and $|1\rangle$. Bring another qubit and the combined systems has four logical states $|00\rangle, |01\rangle, |10\rangle$ and $|11\rangle$. In general n qubits will give us 2^n states representing all possible binary strings of length n. It is important to use subsystems — here qubits — rather than one chunk of matter, for operating on at most n qubits we can reach any of the 2^n states of the composed system. Now, let the qubits interact in a controllable fashion. We are computing. Think about computation as a physical process that evolves a prescribed initial configuration of a computing machine, called `INPUT`, into some final configuration, called `OUTPUT`. We shall refer to the configurations as **states**. Figure 4 shows five consecutive computational steps performed on four distinct states.

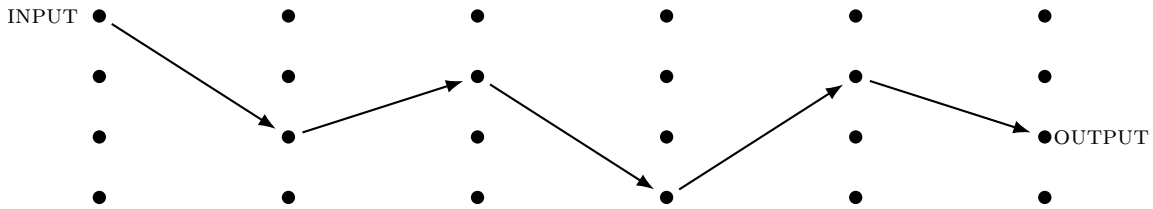


Figure 4. Deterministic computation.

That computation was **deterministic**: every time you run it with the same input, you get the same output. But a computation does not have to be deterministic — we can augment a computing machine by allowing it to "toss an unbiased coin" and to choose its steps randomly. It can then be viewed as a directed tree-like graph where each node corresponds to a state of the machine, and each edge represents one step of the computation, as shown in Figure 5

So we read left to right, and omit the arrowheads.

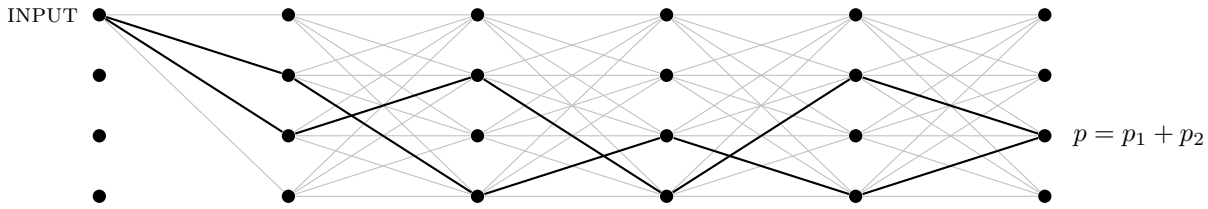


Figure 5. Probabilistic computation.

The computation starts from some initial state (`INPUT`) and it subsequently branches into other nodes representing states reachable with non-zero probability from the initial state. The probability of a particular final state (`OUTPUT`) being reached is equal to the sum of the probabilities along all mutually exclusive paths which connect the initial state with that particular state. Figure 5 shows only two computational paths, but, in general, there could be many more paths (here, up to 256) contributing to the final probability. Quantum computation can be represented by a similar graph, as in 6.

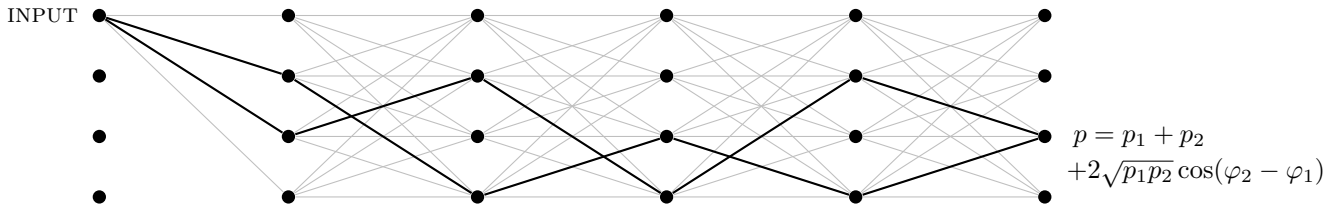


Figure 6. Quantum computation.

For quantum computations, we associate with each edge in the graph the probability *amplitude* that the computation follows that edge. The probability amplitude of a particular path to be followed is the product of amplitudes pertaining to transitions in each step. The probability amplitude of a particular final state being reached is equal to the sum of the amplitudes along all mutually exclusive paths which connect the initial state with that particular state:

$$z = \sum_{\text{all paths } k} z_k.$$

The resulting probability, as we have just seen, is the sum of the probabilities pertaining to each computational path p_k modified by the interference terms:

$$\begin{aligned} p &= |z|^2 \\ &= \sum_{k,j} z_j^\star z_k \\ &= \sum_k p_k + \sum_{k \neq j} \sqrt{p_k p_j}\cos(\varphi_k - \varphi_j). \end{aligned}$$

Quantum computation can be viewed as a complex multi-particle quantum interference involving many computational paths through a computing device. The art of quantum computation is to shape quantum interference, through a sequence of computational steps, enhancing probabilities of correct outputs and suppressing probabilities of the wrong ones.

1.8 Computational complexity

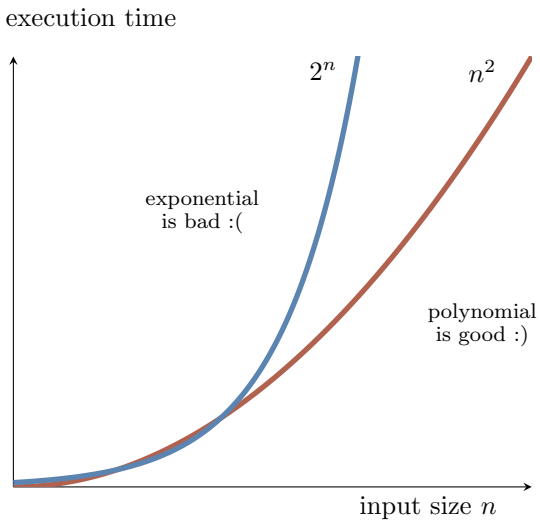
Is there a compelling reason why we should care about quantum computation? It may sound like an extravagant way to compute something that can be computed anyway. Indeed, your standard laptop, given enough time and memory, can simulate pretty much any physical process. In principle, it can also simulate any quantum interference and compute everything that quantum computers can compute. The snag is, this simulation, in general, is very inefficient. And efficiency does matter, especially if you have to wait more than the age of the Universe for your laptop to stop and deliver an answer!

The age of the Universe is currently estimated at 13.772 billion years.

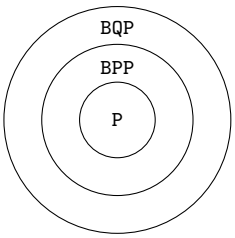
In order to solve a particular problem, computers (classical or quantum) follow a precise set of instructions called an **algorithm**. Computer scientists quantify the efficiency of an algorithm according to how rapidly its running time, or the use of memory, increases when it is given ever larger inputs to work on. An algorithm is said to be **efficient** if the number of elementary operations taken to execute it increases no faster than a polynomial function of the size of the input. We take the input size to be the total number of binary

Note that the technological progress alone, such as increasing the speed of classical computers, will never turn an inefficient algorithm (exponential scaling) into an efficient one (polynomial scaling). Why?

digits (bits) needed to specify the input. For example, using the algorithm taught in elementary school, one can multiply two n digit numbers in a time that grows like the number of digits squared, n^2. In contrast, the fastest-known method for the reverse operation — factoring an n-digit integer into prime numbers — takes a time that grows exponentially, roughly as 2^n. That is considered inefficient.



The class of problems that can be solved by a deterministic computer in polynomial time is represented by the capital letter P, for *polynomial* time. The class of problems that can be solved in polynomial time by a probabilistic computer is called BPP, for *bounded-error probabilistic polynomial* time. It is clear that BPP contains P, since a deterministic computation is a special case of a probabilistic computation in which we never consult the source of randomness. When we run a probabilistic (a.k.a. randomised) computation many times on the same input, we will not get the same answer every time, but the computation is useful if the probability of getting the right answer is high enough. Finally, the complexity class BQP, for *bounded-error quantum polynomial*, is the class of problems that can be solved in polynomial time by a quantum computer.



Since a quantum computer can easily generate random bits and simulate a probabilistic classical computer, BQP certainly contains the class BPP. Here we are interested in problems that are in BQP but not known to be in BPP. The most popular example of such a problem is factoring.

A quantum algorithm, discovered by Peter Shor in 1994, can factor n-digit numbers in a number of steps that grows only as n^2, as opposed to the 2^n that we have classically. Since the intractability of factorisation underpins the security of many methods of encryption, Shor's algorithm was soon hailed as the first 'killer application' for quantum computation: something very useful that only a quantum computer could do. Since then, the hunt has been on for interesting things for quantum computers to do, and at the same time, for the scientific and technological advances that could allow us to build quantum computers.

It must be stressed that not all quantum algorithms are so efficient, in fact many are no faster than their classical counterparts. Which particular problems will lend themselves to quantum speed-ups is an open question.

1.9 Outlook

When the physics of computation was first investigated, starting in the 1960s, one of the main motivations was a fear that quantum-mechanical effects might place fundamental bounds on the accuracy with which physical objects could render the properties of the abstract entities, such as logical variables and operations, that appear in the theory of computation. It turned out, however, that quantum mechanics itself imposes no significant limits, but does break through some of those that classical physics imposed. The quantum world has a richness and intricacy that allows new practical technologies, and new kinds of knowledge. In this course we will merely scratch the surface of the rapidly developing field of quantum computation. We will concentrate mostly on the fundamental issues and skip many experimental details. However, it should be mentioned that quantum computing is a serious possibility for future generations of computing devices. At present it is not clear how and when fully-fledged quantum computers will eventually be built; but this notwithstanding, the quantum theory of computation already plays a much more fundamental role in the scheme of things than its classical predecessor did. I believe that anyone who seeks a fundamental understanding of either physics, computation or logic must incorporate its new insights into their world view.

1.10 Remarks and exercises

1.10.1

As a purely historical remark: back in 1926, Max Born simply postulated the connection between amplitudes and probabilities, but did not get it quite right on his first attempt. In the original paper proposing the probability interpretation of the state vector (wavefunction) he wrote:

Max Born, "Zur Quantenmechanik der StoSSvorgänge", *Zeitschrift für Physik* **37** (1926), 893–867.

> *... If one translates this result into terms of particles only one interpretation is possible.* $\Theta_{\eta,\tau,m}(\alpha,\beta,\gamma)$ *[the wavefunction for the particular problem he is considering] gives the probability* for the electron arriving from the* z *direction to be thrown out into the direction designated by the angles* α,β,γ*...*
>
> * *Addition in proof: More careful considerations show that the probability is proportional to the square of the quantity* $\Theta_{\eta,\tau,m}(\alpha,\beta,\gamma)$.

1.10.2

Suppose that we modified the Born rule, so that probabilities were given by the absolute values of amplitudes raised to power p (for p not necessarily equal to 2). Then admissible physical evolutions must preserve the normalisation of probability. Mathematically speaking, they must be isometries of p-norms.

Recall that the p-norm of vector v, with components $v_1, v_2, \ldots, v_n$, is defined as

$$\sqrt[p]{|v_1|^p + |v_2|^p + \ldots + |v_n|^p}.$$

It is clear that any permutation of vector components and multiplication by phase factors (i.e. unit complex numbers) will leave any p-norm unchanged. It turns out that these complex permutations are the *only* isometries, except for *one* special case! For $p = 2$, the isometries are unitary operations, which form a continuous group; in all other cases we are restricted to discrete permutations. We do not have to go into details of the proof since we can *see* this result.

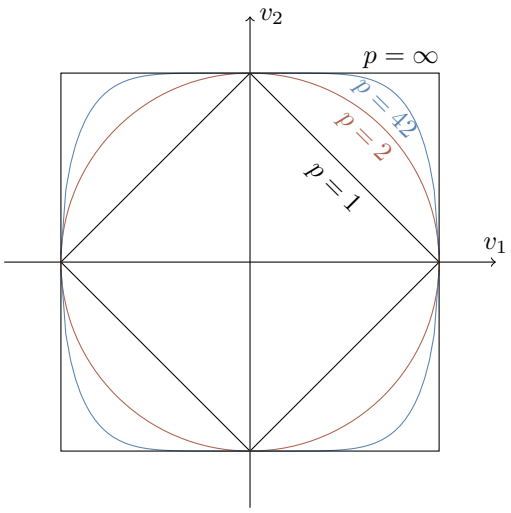


Figure 7. The unit spheres in the p-norm for $p = 1, 2, 42, \infty$.

In particular, the image of the unit sphere must be preserved under probability preserving operations. As we can see in Figure 7, the 2-norm is special because of its rotational invariance — the probability measure picks out no preferred basis in the space of state vectors. Moreover, it respects unitary operations and does not restrict them in any way. If the admissible physical evolutions were restricted to discrete symmetries, e.g. permutations, then there would be no continuity, and no concept of "time" as we know it.

1.10.3

Complex numbers have many applications in physics, however, not until the advent of quantum theory was their ubiquitous and fundamental role in the description of the actual physical world so evident. Even today, their profound link with probabilities appears to be a rather mysterious connection. Mathematically speaking, the set of complex numbers is a field. This is an important algebraic structure used in almost all branches of mathematics. You do not have to know much about algebraic fields to follow these lectures, but still, you should know the basics. Look them up.

a. The set of rational numbers and the set of real numbers are both fields, but the set of integers is not. Why?
b. What does it mean to say that the field of complex numbers is **algebraically closed**?
c. Evaluate each of the following quantities:

$$1 + e^{-i\pi}, \quad |1 + i|, \quad (1 + i)^{42}, \quad \sqrt{i}, \quad 2^i, \quad i^i.$$

d. Here is a simple proof that $+1 = -1$:

$$1 = \sqrt{1} = \sqrt{(-1)(-1)} = \sqrt{-1}\sqrt{-1} = i^2 = -1.$$

What is wrong with it?

1.10.4

A quantum computer starts calculations in some initial state, then follows n different computational paths which lead to the final output. The computational paths are followed with

probability amplitudes $\frac{1}{\sqrt{n}}e^{ik\varphi}$, where φ is a fixed angle $0 < \varphi < 2\pi$ and $k = 0, 1, ...n-1$. Show that the probability of generating the output is

$1 + z + z^2 + \ldots + z^n = \frac{1-z^{n+1}}{1-z}$

$$\frac{1}{n}\left|\frac{1-e^{in\varphi}}{1-e^{i\varphi}}\right|^2 = \frac{1}{n}\frac{\sin^2(n\frac{\varphi}{2})}{\sin^2(\frac{\varphi}{2})}.$$

for $0 < \varphi < 2\pi$, and 1 for $\varphi = 0$. Plot the probability as a function of φ.

1.10.5

Imagine two distant stars, A and B, that emit *identical* photons. If you point a single detector towards them you will register a click every now and then, but you never know which star the photon came from. Now prepare two detectors and point them towards the stars. Assume the photons arrive with the probability amplitudes specified in Figure 8. Every now and then you will register a coincidence: the two detectors will fire.

a. Calculate the probability of a coincidence.
b. Now, assume that $z \approx \frac{1}{r}e^{i\frac{2r\pi}{\lambda}}$, where r is the distance between detectors and the stars. How can we use this to measure r?

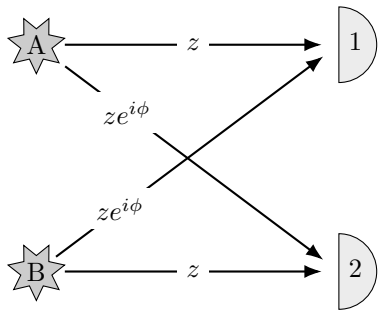


Figure 8. Two photon detectors pointing at two stars, with the probabilities of detection.

1.10.6 Physics against logic?

Now that we have poked our heads into the quantum world, let us see how quantum interference challenges conventional logic and leads to qualitatively different computations. Consider the following task (which we will return to a few more times in later chapters): design a logic gate that operates on a single bit such that, when it is followed by another, identical, logic gate, the output is *always* the negation of the input. Let us call this logic gate **the square root of** `NOT`, or $\sqrt{\texttt{NOT}}$. A simple check, such as an attempt to construct a truth table, should persuade you that there is no such operation in logic. It may seem reasonable to argue that since there is no such operation in logic, $\sqrt{\texttt{NOT}}$ is impossible. Think again!

Figure 9 shows a simple computation, two identical computational steps performed on two states labelled as 0 and 1, i.e. on one bit. An interplay of constructive and destructive interference makes some transitions impossible and the result is the logical `NOT`. Thus, quantum theory declares, the square root of `NOT` is possible. And it does exist! Experimental physicists routinely construct this and many other "impossible" gates in their laboratories. They are the building blocks of a quantum computer. Quantum theory explains the behaviour of $\sqrt{\texttt{NOT}}$, hence, reassured by the physical experiments that corroborate this theory, logicians are now entitled to propose a new logical operation $\sqrt{\texttt{NOT}}$. Why? Because a faithful physical model for it exists in nature.

Write a 2×2 matrix which describes the $\sqrt{\texttt{NOT}}$ operation. Is there just one such a matrix? Suppose you are given a supply of Hadamard and phase gates with tuneable phase settings. How would you construct the $\sqrt{\texttt{NOT}}$ gate?

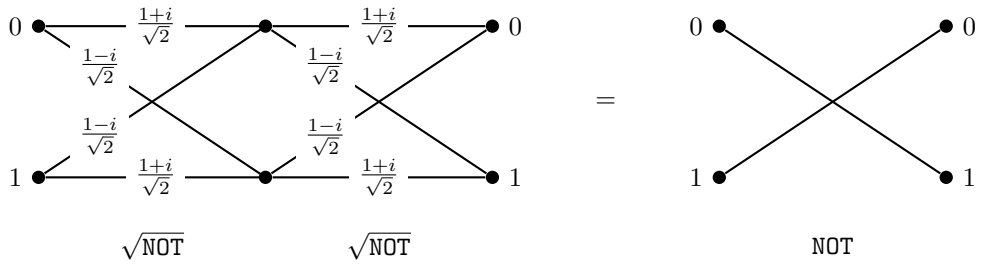


Figure 9. A computation that, when repeated, gives exactly NOT. An unlabelled line means that it has probability 1, and the lack of a line corresponds to having probability 0.

1.10.7 Quantum bomb tester

You have been drafted by the government to help in the demining effort in a former warzone. In particular, retreating forces have left very sensitive bombs in some of the sealed rooms. The bombs are configured such that if even one photon of light is absorbed by the fuse (i.e. if someone looks into the room), the bomb will go off. Each room has an input and output port which can be hooked up to external devices. An empty room will let light go from the input to the output ports unaffected, whilst a room with a bomb will explode if light is shone into the input port and the bomb absorbs even just one photon — see Figure 10.

This is a slightly modified version of a bomb testing problem described by Avshalom Elitzur and Lev Vaidman in *Quantum-mechanical interaction-free measurement*, Found. Phys. **47** (1993), 987-997.

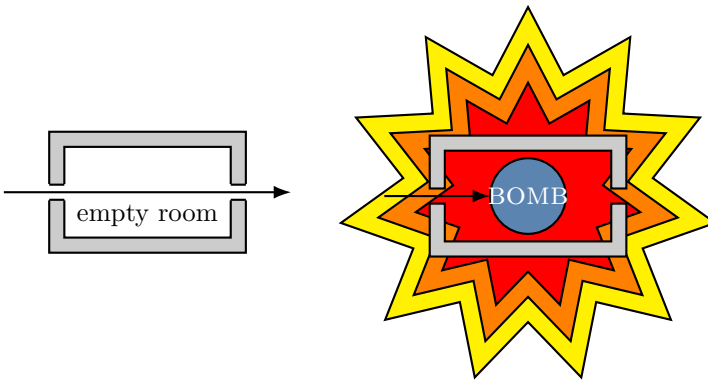


Figure 10. *Left* — the passage of a photon through an empty room. *Right* — the passage of a photon through a room containing a bomb.

Your task is to find a way of determining whether a room has a bomb in it without blowing it up, so that specialised (limited and expensive) equipment can be devoted to defusing that particular room. You would like to know with certainty whether a particular room had a bomb in it.

1. To start with, consider the setup in Figure 11, where the input and output ports are hooked up in the lower arm of a Mach–Zehnder interferometer.

   a. Assume an empty room. Send a photon to input port $|0\rangle$. Which detector, at the output port, will register the photon?
   b. Now assume that the room does contain a bomb. Again, send a photon to input port $|0\rangle$. Which detector will register the photon and with which probability?

Read about Mach–Zehnder interferometers in Chapter 3.

c. Design a scheme that allows you — at least some of the time — to decide whether a room has a bomb in it without blowing it up. If you iterate the procedure, what is its overall success rate for the detection of a bomb without blowing it up?

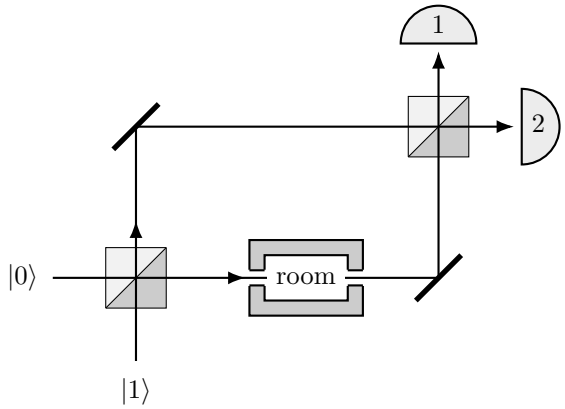


Figure 11. The Mach–Zehnder interferometer hooked up to the bomb-testing room.

2. Assume that the two beam splitters in the interferometer are different. Say the first beam-splitter reflects incoming light with probability r and transmits with probability $t = 1-r$, and the second one transmits with probability r and reflects with probability t. Would the new setup improve the overall success rate of the detection of a bomb without blowing it up?

3. There exists a scheme, involving many beam-splitters and something called the **quantum Zeno effect**, such that the success rate for detecting a bomb without blowing it up approaches 100%. Try to work it out, or find a solution on the internet.

1.10.8

A quantum machine has N perfectly distinguishable configurations. What is the maximum number of computational paths connecting a specific input with a specific output after k steps of the machine? Suppose you are using your laptop to add together amplitudes pertaining to each of the paths. As k and N increase you may need more time and more memory to complete the task. How does the execution time and the memory requirements grow with k and N? Will you need more time or more memory or both?

1.10.9

The classical theory of computation is essentially the theory of the universal Turing machine — the most popular mathematical model of classical computation. Its significance relies on the fact that, given a large but finite amount of time, the universal Turing machine is capable of any computation that can be done by any modern classical digital computer, no matter how powerful. The concept of Turing machines may be modified to incorporate quantum computation, but we will not follow this path. It is much easier to explain the essence of quantum computation talking about quantum logic gates and quantum Boolean networks or circuits. The two approaches are computationally equivalent, even though certain theoretical concepts, e.g. in computational complexity, are easier to formulate precisely using the Turing machine model. The main advantage of quantum circuits is that they relate far more directly to proposed experimental realisations of quantum computation.

1.10.10

In computational complexity the basic distinction is between polynomial and exponential algorithms. Polynomial growth is good and exponential growth is bad, especially if you have to pay for it. There is an old story about the legendary inventor of chess who asked the Persian king to be paid only by a grain of cereal, doubled on each of the 64 squares of a chess board. The king placed one grain of rice on the first square, two on the second, four on the third, and he was supposed to keep on doubling until the board was full. The last square would then have $2^{63} = 9,223,372,036,854,775,808$ grains of rice, more than has been ever harvested on planet Earth, to which we must add the grains of all previous squares, making the total number about twice as large. If we placed that many grains in an unbroken line we would reach the nearest star Alpha Centauri, our closest celestial neighbour beyond the solar system, about 4.4 light-years away. The moral of the story: if whatever you do requires an exponential use of resources, you are in trouble.

One light year (the distance that light travels through a vacuum in one year) is 9.4607×10^{15} metres.

1.10.11

In order to make qualitative distinctions between how different functions grow we will often use the asymptotic big-O notation. For example, suppose an algorithm running on input of size n takes $an^2 + bn + c$ elementary steps, for some positive constants a, b and c. These constants depend mainly on the details of the implementation and the choice of elementary steps. What we really care about is that, for large n, the whole expression is dominated by its quadratic term. We then say that the running time of this algorithm grows as n^2, and we write it as $O(n^2)$, ignoring the less significant terms and the constant coefficients. More precisely, let $f(n)$ and $g(n)$ be functions from positive integers to positive reals. You may think of $f(n)$ and $g(n)$ as the running times of two algorithms on inputs of size n. We say $f = O(g)$, which means that f grows no faster than g, if there is a constant $c > 0$ such that $f(n) \leqslant cg(n)$ for all sufficiently large values of n. Essentially, $f = O(g)$ is a very loose analogue of $f \leqslant g$. In addition to the big-O notation, computer scientists often use Ω for lower bounds: $f = \Omega(g)$ means $g = O(f)$. Again, this is a very loose analogue of $f \geqslant g$.

$f = O(g)$ is pronounced as "f is big-oh of g".

1. When we say that $f(n) = O(\log n)$, why don't we have to specify the base of the logarithm?
2. Let $f(n) = 5n^3 + 1000n + 50$. Is $f(n) = O(n^3)$, or $O(n^4)$, or both?
3. Which of the following statements are true?
   a. $n^k = O(2^n)$ for any constant k
   b. $n! = O(n^n)$
   c. if $f_1 = O(g)$ and $f_2 = O(g)$, then $f_1 + f_2 = O(g)$.

1.10.12

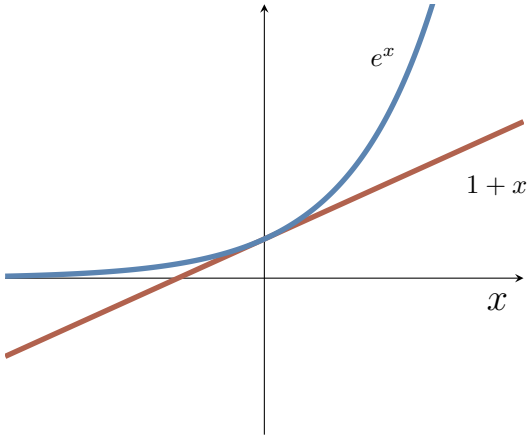
There exists a randomised algorithm which tests whether a given number N is prime. The algorithm always returns `yes` when N is prime, and the probability it returns `yes` when N is not prime is ϵ, which not greater than a half (independently, each time you run the algorithm). You run this algorithm (for the same N) r times and each time the algorithm returns `yes`. What is the probability that N is not prime?

Primality used to be given as the classic example of a problem in BPP but not P. However, in 2002 a deterministic polynomial time test for primality was proposed by Manindra Agrawal, Neeraj Kayal, and Nitin Saxena. Thus, since 2002, primality has been in P.

1.10.13

Suppose a randomised algorithm solves a decision problem, returning `yes` or `no` answers. It gets the answer wrong with a probability not greater than $\frac{1}{2} - \delta$, where $\delta > 0$ is a constant.. If we are willing to accept a probability of error no larger than ϵ, then it suffices to run the computation r times, where $r = O(\log 1/\epsilon)$.}

This result is known as the **Chernoff bound**.



1. If we perform this computation r times, how many possible sequences of outcomes are there?
2. Give a bound on the probability of any particular sequence with w wrong answers.
3. If we look at the set of r outcomes, we will determine the final outcome by performing a majority vote. This can only go wrong if $w > r/2$. Give an upper bound on the probability of any single sequence that would lead us to the wrong conclusion.
4. Using the bound $1 - x \leqslant e^{-x}$, conclude that the probability of our coming to the wrong conclusion is upper bounded by $e^{-2r\delta^2}$.

2 Qubits

*About **quantum bits** and **quantum circuits**, including the "impossible" **square root of NOT**, as well as an introduction to **single-qubit unitaries** and rotations of the **Bloch sphere**, and the implications concerning **universal gates**.*

When studying classical information theory, one single bit isn't usually the most interesting object to think about — it's either 0 or 1. Yet in the quantum case, just working with one "quantum bit" (which we call a **qubit**) opens up a whole world of interesting mathematics. In fact, **single qubit interference** is arguably one of the fundamental building blocks for quantum computing, and deserves to be thoroughly investigated and understood.

2.1 Composing quantum operations

In order to understand something in its full complexity it is always good to start with the simplest case. Let us take a closer look at quantum interference in the simplest possible computing machine: the one that has only two distinguishable configurations — two quantum states — which we label as $|0\rangle$ and $|1\rangle$. We prepare the machine in some input state, usually $|0\rangle$, and let it **evolve**: the machine undergoes a prescribed sequence of computational steps, each of which induces transitions between the two "computational states", $|0\rangle$ and $|1\rangle$. The machine then ends in the output state $|\psi\rangle = \alpha_0|0\rangle + \alpha_1|1\rangle$, meaning the two outputs, $|0\rangle$ and $|1\rangle$, are reached with probability amplitudes α_0 and α_1, respectively. In the process of computation each computational step U (also referred to as an **operation**) sends state $|k\rangle$ to state $|l\rangle$, where $k, l = 0, 1$, but only with some **amplitude** U_{lk}. We write this as

$$|k\rangle \longmapsto \sum_l U_{lk}|l\rangle.$$

(watch out for the order of the indices).

Thus any computational step U of this machine can be described by a matrix which tabulates all the transition amplitudes:

$$U = \begin{bmatrix} U_{00} & U_{01} \\ U_{10} & U_{11} \end{bmatrix}.$$

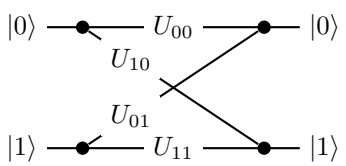
The matrix element U_{lk} represents the amplitude of transition from state $|k\rangle$ to state $|l\rangle$ (again, watch the order of indices). To be clear, the entries in this matrix are not any random complex numbers: their moduli squared represent transition probabilities, which in turn implies that such matrices must be **unitary**.

Recall that matrix U is called **unitary** if

$$U^\dagger U = UU^\dagger = \mathbf{1}$$

where the **adjoint** or **Hermitian conjugate** $U^\dagger$ of any matrix U with complex entries U_{ij} is obtained by taking the complex conjugate of every element in the matrix and then interchanging rows and columns ($U^\dagger_{kl} = U^\star_{lk}$).

We can also describe U by drawing a diagram, which contains exactly the same information as the matrix representation, but just in a different form:



Now how can we find some quantum interference to study? Consider two computational steps, U and V. What is the amplitude that input $|k\rangle$ will generate output $|m\rangle$? We

have to check all computational paths leading from input $|k\rangle$ to output $|m\rangle$ and add the corresponding amplitudes. For example, as you can see in Figure 12, input $|0\rangle$ and output $|1\rangle$ are connected by the two computational paths: $|0\rangle \mapsto |0\rangle \mapsto |1\rangle$ (amplitude $V_{10}U_{00}$) and $|0\rangle \mapsto |1\rangle \mapsto |1\rangle$ (amplitude $V_{11}U_{10}$). Thus the total amplitude that input $|0\rangle$ gives output $|1\rangle$ is the sum $V_{10}U_{00} + V_{11}U_{10}$, and when we take the modulus squared of this expression we will see the interference term.

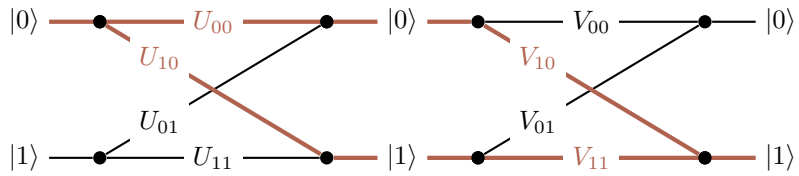


Figure 12. The composition of two computational steps, U and V, with the possible paths from $|0\rangle$ to $|1\rangle$ highlighted.

In general, given U and V

$$|k\rangle \longmapsto \sum_l U_{lk}|l\rangle$$
$$|l\rangle \longmapsto \sum_m V_{ml}|m\rangle$$

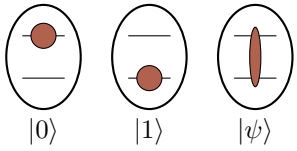
we can compose the two operations: we first apply U, and then V, to obtain

$$\begin{aligned}|k\rangle &\longmapsto \sum_l U_{lk}\left(\sum_m V_{ml}|m\rangle\right)\\ &= \sum_m \left(\sum_l V_{ml}U_{lk}\right)|m\rangle\\ &= \sum_m (VU)_{mk}|m\rangle.\end{aligned}$$

If you want to hone your quantum intuition think about it the following way. The amplitude that input $|k\rangle$ evolves to $|m\rangle$ via a specific intermediate state $|l\rangle$ is given by $V_{ml}U_{lk}$ (evolutions are independent so the amplitudes are multiplied). This done, we have to sum over all possible values of l (the transition can occur in several mutually exclusive ways so the amplitudes are added) to obtain $\sum_l V_{ml}U_{lk}$. Thus the matrix multiplication VU (watch the order of matrices) in one swoop takes care of multiplication and addition of amplitudes corresponding to different computational paths.

2.2 Quantum bits, called "qubits"

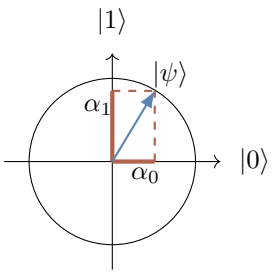
A two-state machine that we have just described in abstract terms is usually realised as a controlled evolution of a two state system, called a quantum bit or a qubit. For example, state $|0\rangle$ may be chosen to be the lowest energy state of an atom (the **ground state**), and state $|1\rangle$ a higher energy state (the **excited state**). Pulses of light of appropriate frequency, duration, and intensity can take the atom back and forth between the basis states $|0\rangle$ and $|1\rangle$ (implementing logical `NOT`).



Some other pulses (say, half the duration or intensity) will take the atom into states that have no classical analogue. Such states are called **coherent superpositions** of $|0\rangle$ and $|1\rangle$, and represent a qubit in state $|0\rangle$ with some amplitude α_0 and in state $|1\rangle$ with some other amplitude α_1. This is conveniently represented by a state vector

$$|\psi\rangle = \alpha_0|0\rangle + \alpha_1|1\rangle \leftrightarrow \begin{bmatrix}\alpha_0\\ \alpha_1\end{bmatrix}$$

which we can also draw graphically:



A **qubit** is a quantum system in which the Boolean states 0 and 1 are represented by a prescribed pair of normalised and mutually orthogonal quantum states labelled by $|0\rangle$ and $|1\rangle$. The two states form a so-called **computational** (or **standard**) basis, and so any other state of an isolated qubit can be written as a coherent superposition

$$|\psi\rangle = \alpha_0|0\rangle + \alpha_1|1\rangle$$

for some α_0 and α_1 such that $|\alpha_0|^2 + |\alpha_1|^2 = 1$.
In practice, a qubit is typically a microscopic system, such as an atom, a nuclear spin, or a polarised photon.

As we have already mentioned, any computational step, that is, any physically admissible operation U on a qubit, is described by a (2×2) unitary matrix U. It modifies the state of the qubit as

$$|\psi\rangle \longmapsto |\psi'\rangle = U|\psi\rangle$$

which we can write explicitly as

$$\begin{bmatrix}\alpha'_0\\ \alpha'_1\end{bmatrix} = \begin{bmatrix}U_{00} & U_{01}\\ U_{10} & U_{11}\end{bmatrix}\begin{bmatrix}\alpha_0\\ \alpha_1\end{bmatrix}$$

That is, the operation U turns the state $|\psi\rangle$, with components α_k, into the state $|\psi'\rangle = U|\psi\rangle$, with components $\alpha'_l = \sum_k U_{lk}\alpha_k$.

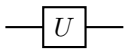
Here we are talking about *isolated* systems. As you will soon learn, a larger class of physically admissible operations is described by completely positive maps. It may sound awfully complicated but, as you will soon see, it is actually very simple.

2.3 Quantum gates and circuits

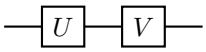
Atoms, trapped ions, molecules, nuclear spins and many other quantum objects, which we call qubits, can be used to implement simple quantum interference, and hence simple quantum computation. There is no need to learn about physics behind these diverse technologies if all you want is to understand the basics of quantum computation. We may now conveniently forget about any specific experimental realisation of a qubit and just remember that any manipulations on qubits have to be performed by physically admissible operations, and that such operations are represented by unitary transformations.

A **quantum (logic) gate** is a device which performs a fixed unitary operation on selected qubits in a fixed period of time, and a **quantum circuit** is a device consisting of quantum logic gates whose computational steps are synchronised in time. The **sizes** of the circuit is the number of gates it contains.

Some unitary U acting on a single qubit is represented diagrammatically as



This diagram should be read from left to right. The horizontal line represents a qubit that is inertly carried from one quantum operation to another. We often call this line a **quantum wire**. The wire may describe translation in space (e.g. atoms travelling through cavities) or translation in time (e.g. a sequence of operations performed on a trapped ion). A sequence of two gates acting on the same qubit, say U followed by V, is represented by



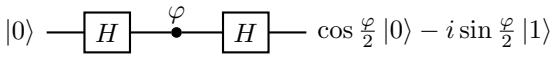
and is described by the matrix product VU (note the order in which we multiply the matrices).

2.4 Single qubit interference

Let me now describe what is probably the most important sequence of operations performed on a single qubit, namely a generic **single qubit interference**. It is typically constructed as a sequence of three elementary operations:

1. the Hadamard gate
2. a phase-shift gate
3. the Hadamard gate again.

We represent it graphically as



Hadamard	$H = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$	$\begin{aligned} \|0\rangle &\longmapsto \tfrac{1}{\sqrt{2}}(\|0\rangle + \|1\rangle) \\ \|1\rangle &\longmapsto \tfrac{1}{\sqrt{2}}(\|0\rangle - \|1\rangle) \end{aligned}$
Phase-shift	$P_\varphi = \begin{bmatrix} 1 & 0 \\ 0 & e^{i\varphi} \end{bmatrix}$	$\begin{aligned} \|0\rangle &\longmapsto \|0\rangle \\ \|1\rangle &\longmapsto e^{i\varphi}\|1\rangle \end{aligned}$

You will see it over and over again, for it is quantum interference that gives quantum computation additional capabilities.

Indeed, you have already seen this sequence: recall our study of Ramsey interferometry, and note how this is essentially the same!

Something that many explanations of quantum computing say is the following: "quantum computers are quicker because they evaluate all possible solutions at once, in parallel". **This is not accurate.**
Firstly, quantum computers are not necessarily "quicker" than classical computers, but can simply implement quantum algorithms, some of which *are* quicker than their classical counterparts. Secondly, the idea that they "just do all the possible computations at once" is false — instead, they rely on thoughtfully using interference (which can be constructive or destructive) to modify the probabilities of specific outcomes.
The power of quantum computing comes from quantum interference.

The product of the three matrices $HP_\varphi H$ describes the action of the whole circuit: it gives the transition amplitudes between states $|0\rangle$ and $|1\rangle$ at the input and the output as

$$e^{i\frac{\varphi}{2}}\begin{bmatrix} \cos\varphi/2 & -i\sin\varphi/2 \\ -i\sin\varphi/2 & \cos\varphi/2 \end{bmatrix} = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & e^{i\varphi} \end{bmatrix}\frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

Given that our input state is almost always $|0\rangle$, it is sometimes much easier and more instructive to step through the execution of this circuit and follow the evolving state. The interference circuit effects the following sequence of transformations:

We ignore the global phase factor $e^{i\frac{\varphi}{2}}$.

$$\begin{aligned} |0\rangle &\xmapsto{H} \frac{1}{\sqrt{2}}\left(|0\rangle + |1\rangle\right) \\ &\xmapsto{P_\phi} \frac{1}{\sqrt{2}}\left(|0\rangle + e^{i\phi}|1\rangle\right) \\ &\xmapsto{H} \cos\frac{\phi}{2}|0\rangle - i\sin\frac{\phi}{2}|1\rangle. \end{aligned}$$

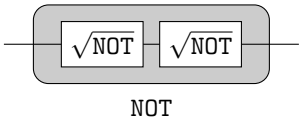
The first Hadamard gate prepares an equally weighted superposition of $|0\rangle$ and $|1\rangle$ and the second one closes the interference by bringing the interfering paths together. The phase shift φ effectively controls the evolution and determines the output. The probabilities of finding the qubit in state $|0\rangle$ or $|1\rangle$ at the output are, respectively,

$$\begin{aligned} \Pr(0) &= \cos^2\frac{\phi}{2} \\ \Pr(1) &= \sin^2\frac{\phi}{2}. \end{aligned}$$

This simple quantum process contains, in a nutshell, the essential ingredients of quantum computation. This sequence (Hadamard–phase shift–Hadamard) will appear over and over again. It reflects a natural progression of quantum computation: first we prepare different computational paths, then we evaluate a function which effectively introduces phase shifts into different computational paths, then we bring the computational paths together at the output.

2.5 The square root of NOT

Now that we have poked our heads into the quantum world, let us see how quantum interference challenges conventional logic. Consider a following task: design a logic gate that operates on a single bit and such that when it is followed by another, identical, logic gate the output is always the negation of the input. Let us call this logic gate **the square root of** NOT, or $\sqrt{\texttt{NOT}}$.



A simple check, such as an attempt to construct a truth table, should persuade you that there is no such operation in logic. It may seem reasonable to argue that since there is no such operation in logic, $\sqrt{\texttt{NOT}}$ is impossible. But it does exist! Experimental physicists routinely construct such "impossible" gates in their laboratories. It is a physically admissible operation described by the unitary matrix

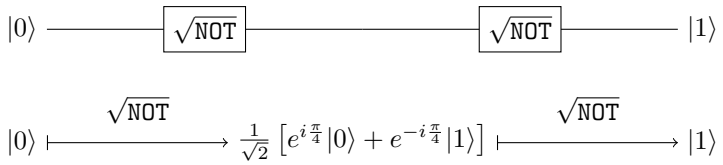
There are infinitely many unitary operations that act as the square root of NOT.

$$\sqrt{\texttt{NOT}} = \frac{1}{2}\begin{bmatrix} 1+i & 1-i \\ 1-i & 1+i \end{bmatrix} = \frac{1}{\sqrt{2}}\begin{bmatrix} e^{i\frac{\pi}{4}} & e^{-i\frac{\pi}{4}} \\ e^{-i\frac{\pi}{4}} & e^{i\frac{\pi}{4}} \end{bmatrix}.$$

Indeed,

$$\frac{1}{2}\begin{bmatrix} 1+i & 1-i \\ 1-i & 1+i \end{bmatrix}\frac{1}{2}\begin{bmatrix} 1+i & 1-i \\ 1-i & 1+i \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

We could also step through the circuit diagram and follow the evolution of the state vector:



Or, if you prefer to work with column vectors and matrices, you can write the two consecutive application of $\sqrt{\texttt{NOT}}$ to state $|0\rangle$ as

$$\begin{bmatrix} 0 \\ 1 \end{bmatrix} \longleftarrow\!\shortmid \frac{1}{\sqrt{2}}\begin{bmatrix} e^{i\frac{\pi}{4}} \\ e^{-i\frac{\pi}{4}} \end{bmatrix} \longleftarrow\!\shortmid \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

(following a well established convention, the above should be read from *right to left*), where each $\longleftarrow\!\shortmid$ denotes multiplication by $\frac{1}{\sqrt{2}}\begin{bmatrix} e^{i\frac{\pi}{4}} & e^{-i\frac{\pi}{4}} \\ e^{-i\frac{\pi}{4}} & e^{i\frac{\pi}{4}} \end{bmatrix}$.

Just remember that circuits diagrams are read from *left to right*, and vector and matrix operations go from *right to left*.

One way or another, quantum theory explains the behaviour of $\sqrt{\texttt{NOT}}$, and so, reassured by the physical experiments that corroborate this theory, logicians are now entitled to propose a new logical operation $\sqrt{\texttt{NOT}}$. Why? Because a faithful physical model for it exists in nature!

We discuss this in more detail in [Appendix: Physics against logic, via beamsplitters].

2.6 Phase gates galore

As well as the generic phase gate P_φ, let us mention three specific phase gates that will frequently pop up (two of which have rather confusing names, at first glance!).

Generic phase-shift	$P_\varphi = \begin{bmatrix} 1 & 0 \\ 0 & e^{i\varphi} \end{bmatrix}$	$\begin{aligned} \lvert 0\rangle &\longmapsto \lvert 0\rangle \\ \lvert 1\rangle &\longmapsto e^{i\varphi}\lvert 1\rangle \end{aligned}$
Phase-flip	$Z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$	$\begin{aligned} \lvert 0\rangle &\longmapsto \lvert 0\rangle \\ \lvert 1\rangle &\longmapsto -\lvert 1\rangle \end{aligned}$
$\pi/4$**-phase**	$S = \begin{bmatrix} 1 & 0 \\ 0 & i \end{bmatrix}$	$\begin{aligned} \lvert 0\rangle &\longmapsto \lvert 0\rangle \\ \lvert 1\rangle &\longmapsto i\lvert 1\rangle \end{aligned}$
$\pi/8$**-phase**	$T = \begin{bmatrix} 1 & 0 \\ 0 & e^{i\frac{\pi}{4}} \end{bmatrix}$	$\begin{aligned} \lvert 0\rangle &\longmapsto \lvert 0\rangle \\ \lvert 1\rangle &\longmapsto e^{i\frac{\pi}{4}}\lvert 1\rangle \end{aligned}$

Note that the phase gate P_φ is only defined up to a global phase factor, and so we can write its matrix either as

$$P_\varphi = \begin{bmatrix} 1 & 0 \\ 0 & e^{i\varphi} \end{bmatrix}$$

or as

$$P_\varphi = \begin{bmatrix} e^{-i\frac{\varphi}{2}} & 0 \\ 0 & e^{i\frac{\varphi}{2}} \end{bmatrix}$$

In general, states differing only by a global phase are physically indistinguishable, and so it is physical experimentation that leads us to this mathematical choice of only defining things up to a global phase.

The first version is more common in the quantum information science community, but the second one is sometimes more convenient to use, as it has determinant 1, and hence belongs to the group $\mathrm{SU}(2)$. We will occasionally switch to the $\mathrm{SU}(2)$ version of a phase gates, and this is where the $\pi/4$-phase and $\pi/8$-phase gates get their names, since their $\mathrm{SU}(2)$ versions have $e^{\mp i\pi/4}$ and $e^{\mp i\pi/8}$ (respectively) on the diagonal.

The remaining gate (Z) is arguably the most important specific phase gate, since it is one of the **Pauli operators**, which we will now discuss.

2.7 Pauli operators

Adding to our collection of common single-qubit gates, we now look at the three **Pauli operators** σ_x, σ_y, and σ_z, also denoted by X, Y, and Z (respectively). These three operators, combined with the identity, satisfy a lot of nice formal properties, which we shall examine briefly here, and then return to in more detail in Chapter 3.

We use the standard basis $\{|0\rangle, |1\rangle\}$ most of the time, and so often refer to operators as matrices.

Identity	$\mathbf{1} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$	$\begin{aligned} \|0\rangle &\longmapsto \|0\rangle \\ \|1\rangle &\longmapsto \|1\rangle \end{aligned}$
Bit-flip	$X = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$	$\begin{aligned} \|0\rangle &\longmapsto \|1\rangle \\ \|1\rangle &\longmapsto \|0\rangle \end{aligned}$
Bit-phase-flip	$Y = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}$	$\begin{aligned} \|0\rangle &\longmapsto i\|1\rangle \\ \|1\rangle &\longmapsto -i\|0\rangle \end{aligned}$
Phase-flip	$Z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$	$\begin{aligned} \|0\rangle &\longmapsto \|0\rangle \\ \|1\rangle &\longmapsto -\|1\rangle \end{aligned}$

The identity is just a quantum wire, and we have already seen the X and Z gates in Phase gates galore, as the bit-flip and phase-flip (respectively). Note that, of these latter two, only the X gate has a classical analogue (as the logical `NOT` operator). The remaining gate, the Y operator, describes the combined effect of both the bit- and the phase-flip: $ZX = iY$.

In fact, this is just one of the equations that the Pauli matrices satisfy. The Pauli matrices are unitary and Hermitian, they square to the identity, and they anti-commute. By this last point, we mean that

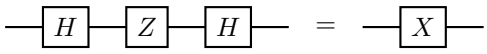
$$\begin{aligned} XY + YX &= 0, \\ XZ + ZX &= 0, \\ YZ + ZY &= 0. \end{aligned}$$

As already mentioned, they satisfy $ZX = iY$, but also any cyclic permutation of this equation.

These operators are also called **sigma matrices**, or **Pauli spin matrices**. They are so ubiquitous in quantum physics that they should certainly be memorised.

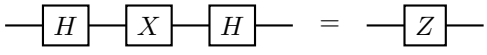
2.8 From bit-flips to phase-flips, and back again

The Pauli Z gate is a special case of a phase gate P_φ with $\varphi = \pi$. When we insert it into the interference circuit we obtain



If you wish to verify this, write the Hadamard gate as $H = (X + Z)/\sqrt{2}$ and use the properties of the Pauli operators. So the Hadamard gate turns phase-flips into bit-flips, but it also turns bit-flips into phase-flips:

$HXH = Z$
$HZH = X$
$HYH = -Y$



Let us also add, for completeness, that $HYH = -Y$. You will see these identities again and again, especially when we discuss quantum error corrections.

2.9 Any unitary operation on a single qubit

There are infinitely many unitary operations that can be performed on a single qubit. In general, any complex $(n \times n)$ matrix has n^2 complex entries, and can thus be specified by

Unitaries, such as H, that take the three Pauli operators to the Pauli operators via conjugation form the so-called **Clifford group**, which we will meet later on. Which phase gate is in the Clifford group of a single qubit?

$2n^2$ real independent parameters. The unitarity constraint removes n^2 of these, and so any unitary $(n \times n)$ matrix has n^2 real independent parameters. In particular, we need *four* real parameters to specify a (2×2) unitary matrix. If we are prepared to ignore global phase factors (which we are) then there are only three real parameters left. So, with this in mind, can we construct and implement any unitary on a single qubit in some simple way?

Yes, we can.

Any unitary operation on a qubit (up to an overall multiplicative phase factor) can be implemented by a circuit containing just two Hadamards and three phase gates, with adjustable phase settings, as in Figure 13.

Figure 13. The universal circuit for unitary (2×2) matrices.

If we multiply the matrices corresponding to each gate in the network (remember that the order of matrix multiplication is reversed) we obtain

$$U(\alpha, \beta, \gamma) = \begin{bmatrix} e^{-i\left(\frac{\alpha+\beta}{2}\right)} \cos\varphi/2 & -ie^{i\left(\frac{\alpha-\beta}{2}\right)} \sin\varphi/2 \\ -ie^{-i\left(\frac{\alpha-\beta}{2}\right)} \sin\varphi/2 & e^{i\left(\frac{\alpha+\beta}{2}\right)} \cos\varphi/2 \end{bmatrix}.$$

Any (2×2) unitary matrix (up to global phase) can be expressed in this form using the three independent real parameters, α, β, and φ, which take values in $[0, 2\pi]$. In order to see that this construction does what it claims, let us explore an intriguing mathematical connection between single qubit unitaries and rotations in three dimensions.

2.10 The Bloch sphere

Unitary operations on a single qubit form a group. More precisely, the set of all (2×2) unitary matrices forms a non-abelian group under the matrix multiplication, denoted by $\mathrm{U}(2)$. It turns out that compositions of single qubit unitaries behave pretty much the same as compositions of rotations in three dimensions. Technically speaking, we claim that $\mathrm{U}(2)/\mathrm{U}(1) \cong \mathrm{SO}(3)$. That is, (2×2) unitaries, up to global phase, form a group which is isomorphic to the group of rotations in three dimensions, denoted by $\mathrm{SO}(3)$. This isomorphism helps to visualise the actions of single-qubit gates.

Note that $\mathrm{U}(1) \cong \mathbb{C}^\times$, where $\mathbb{C}^\times$ is the multiplicative group of unit elements of the complex numbers, i.e. the set $\mathbb{C} \setminus \{0\}$ with the group operation given by multiplication.

There are many ways to introduce this isomorphism. Here we will just show how to represent single-qubit state vectors in terms of Euclidean vectors in three dimensions; in Chapter 3 we will actually relate unitary operations on state vectors to rotations in this Euclidean space, demonstrating this isomorphism.

That is, we have the group $\mathrm{U}(2)$ acting on the space of single-qubit state vectors, and we have the group $\mathrm{SO}(3)$ acting on the unit sphere $S^2 \subset \mathbb{R}^3$. In this chapter we will discuss how to go from one *space* (i.e. the thing being acted upon by the group) to the other; in Chapter 3 we will discuss how to go from one *group* (i.e. the thing acting on the space) to the other.

Any single qubit state can be written as $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$, constrained by the relation $|\alpha|^2 + |\beta|^2 = 1$. This suggests a more natural parametrisation as

$$|\psi\rangle = \cos\frac{\theta}{2} e^{i\varphi_0}|0\rangle + \sin\frac{\theta}{2} e^{i\varphi_1}|1\rangle.$$

There is a good reason to use $\theta/2$ instead of θ, which we will explain later on.

We can then factor out a global phase:

$$|\psi\rangle = e^{i\varphi_0}\left(\cos\frac{\theta}{2}|0\rangle + \sin\frac{\theta}{2} e^{i\varphi}|1\rangle\right),$$

and even remove it completely, since states that are identical up to a global phase are physically indistinguishable.

The parametrisation in terms of θ and φ should remind you of spherical polar coordinates for the surface of a sphere.

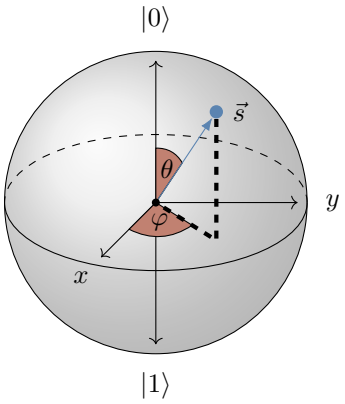


Figure 14. The Bloch sphere, with the point $\vec{s}$ corresponding to $|\psi\rangle$ marked.

We call this sphere the **Bloch sphere**, and the unit vector $\vec{s}$ defined by θ and φ the **Bloch vector**. This is a very useful way to visualise quantum states of a single qubit and unitary operations that we perform on it. Any unitary action on the state vector will induce a rotation of the corresponding Bloch vector. But what kind of rotation?

We will revisit this construction again in more detail, and from a slightly different point of view, in Chapter 3.

We give a complete answer to this question in Chapter 3, but we might as well give some specific results here first, since some are easy enough to calculate "by hand". Note that *any two orthogonal state vectors appear on the Bloch sphere as two Bloch vectors pointing in opposite directions*. Now, the two eigenvectors of a single-qubit unitary U must be orthogonal, and thus define an axis running through the centre of the Bloch sphere. This is the axis about which the Bloch vector is rotated when U acts on the corresponding state vector. The rotation angle α is given by the eigenvalues of U, which, up to a global phase factor, are of the form $e^{\mp i\alpha/2}$.

It is instructive to work out few simple cases and get a feel for the rotations corresponding to the most common unitaries. For example, it is easy to check that a phase gate P_α acts by

$$\cos\frac{\theta}{2}|0\rangle + e^{i\varphi}\sin\frac{\theta}{2}|1\rangle \longmapsto \cos\frac{\theta}{2}|0\rangle + e^{i(\varphi+\alpha)}\sin\frac{\theta}{2}|1\rangle.$$

The azimuthal angle changes from φ to $\varphi+\alpha$, and so the Bloch sphere is rotated anticlockwise by α about the z-axis. The Bloch vectors corresponding to the two eigenvectors of P_α, namely $|0\rangle$ and $|1\rangle$, define the axis of the rotation.

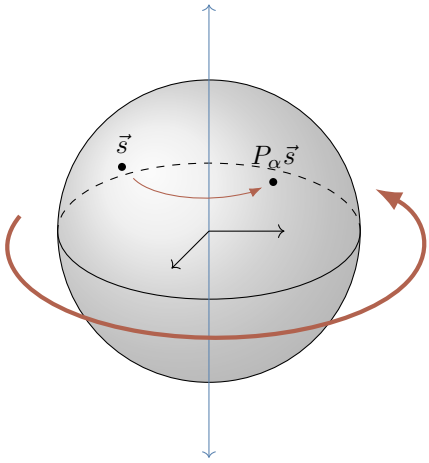


Figure 15. Phase gates P_α represent rotations of the Bloch sphere around the z-axis.

As previously mentioned, the Pauli operator $Z = \sigma_z$ is a special case of a phase gate, and represents rotation by $180°$ (that is, π radians), about the z-axis. You can also verify that $X = \sigma_x$, with eigenvectors $(|0\rangle \pm |1\rangle)/\sqrt{2}$, represents rotation by $180°$ about the x-axis, and $Y = \sigma_y$, with eigenvectors $(|0\rangle \pm i|1\rangle)/\sqrt{2}$, represents rotation by $180°$ about the y-axis.

!!!TODO!!! points on the intersection of the axes with the Bloch sphere are exactly the eigenstates of the corresponding Pauli operator

How about the Hadamard gate? Like the Pauli operators, it squares to the identity ($H^2 = \mathbf{1}$), which implies that its eigenvalues are ± 1. Thus it will correspond to a rotation by $180°$. But about which axis? This time, rather than finding eigenvectors of H, we notice that $HXH = Z$ and $HZH = X$, thus H must swap x- and z-axes, turning rotations about the z-axis into rotations about the x-axis, and vice versa. The Hadamard gate must then represent rotation by $180°$ about the diagonal $(x + z)$-axis. You may also notice that, after this rotation, the y-axis points in the opposite direction, which seems to be related to another identity: $HYH = -Y$. This is not a coincidence.

One can show that the effect of the rotation represented by unitary U on the Bloch vector with components s_x, s_y, s_z is summarised in the formula

Again, see Chapter 3.

$$U(s_x X + s_y Y + s_z Z)U^\dagger = s'_x X + s'_y Y + s'_z Z,$$

where s'_x, s'_y, and s'_z are the components of the rotated Bloch vector.

2.10.1 Drawing points on the Bloch sphere

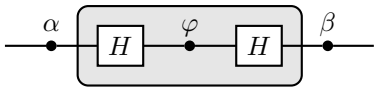
We know that the state $|0\rangle$ corresponds to the north pole of the Bloch sphere, and the state $|1\rangle$ to the south, but what about an arbitrary state $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$? By definition, we can find the parametrisation in terms of θ and φ, but there is also a neat "trick" for finding the point on the Bloch sphere that corresponds to $|\psi\rangle$, which goes as follows.

1. Calculate $\lambda = \beta/\alpha$ (assuming that $\alpha \neq 0$, since otherwise $|\psi\rangle = |1\rangle$).
2. Write $\lambda = \lambda_x + i\lambda_y$ and mark the point $p = (\lambda_x, \lambda_y)$ in the xy-plane (i.e. the plane $\{z = 0\}$).
3. Draw the line going through the south-pole and the point p. This will intersect the Bloch sphere in exactly one other point, and this is exactly the point corresponding to $|\psi\rangle$.

Note that this lets you *draw* the point on the sphere, but doesn't (immediately) give you the *coordinates* for it. That is, this method is nice for geometric visualisation, but the parametrisation method is much better when it comes to actually doing calculations.

2.11 Composition of rotations

We are now in a position understand the circuit in Figure 13 in geometric terms. Recall that *any* rotation in the Euclidean space can be performed as a sequence of three rotations: one about the z-axis, one about the x-axis, and one more about z-axis. The circuit does exactly this:



The first phase gate effects rotation by α about the z-axis, the second phase gate is sandwiched between the two Hadamard gates, and these three gates together effect rotation by φ about the x-axis, and, finally, the third phase gates effects rotation by β about the z-axis. So we can implement any unitary U by choosing the three phase shifts, α, φ, and β, which are known as the three **Euler's angles**.

2.12 A finite set of universal gates

The Hadamard gate and the phase gates, with adjustable phases, allow us to implement an arbitrary single-qubit unitary *exactly*. The tacit assumption here is that we have here *infinitely many* phase gates: one gate for each phase. In fact, we can pick just one phase gate, namely any phase gate P_α with the phase α that is incommensurate with π. It is clear that repeated iteration of P_α can be used to approximate any other phase gate to arbitrary accuracy: indeed, rotate the Bloch sphere by α about the z-axis sufficiently many times and you end up as close as you please to any other rotation about the z-axis.

That is, there do *not* exist any $m, n \in \mathbb{Z}$ such that $m\alpha = n\pi$. For example, it suffices to take α to be rational.

If you want to be ϵ-close to the desired angle of rotation, then you may need to repeat the rotation by α roughly $1/\epsilon$ times. Indeed, within n applications (for $n\alpha \gg 2\pi$) of P_α, we expect the accessible angles to be approximately evenly distributed within the range $[0, 2\pi]$, i.e. any angle of rotation can be achieved to an accuracy of $\epsilon = 2\pi/n$ by using up to $n \approx 1/\epsilon$ applications of P_α. So we can use *just one* phase gate to *approximate* the *three* phase gates in the circuit in Figure 13.

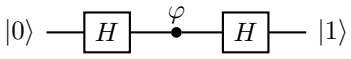
There are other ways of implementing irrational rotations of the Bloch sphere. For example, take the Hadamard gate and the T gate. You can check that the compositions $THTH$ and $HTHT$ represent rotations by angles which are irrational multiples of π, about two different axes. We can then compose a sequence of these two rotations to approximate any other rotation of the sphere. This may look very nice in theory, but there are issues with the actual physical implementation of this approach. All the gates in the circuit will operate with finite precision, and the phase gates will deviate from implementing the required irrational rotations. It turns out, however, that we can tolerate minor imperfections; the final result will not be that far off.

For more details on all the above, see Chapter 3.

2.13 Remarks and Exercises

2.13.1

Consider the usual quantum interference circuit:



Suppose you can control the input of the circuit and measure the output, but you do not know the phase shift φ introduced by the phase gate. You prepare input $|0\rangle$ and register output $|1\rangle$. What can you say about φ?

Now you are promised that φ is either 0 or π. You can run the circuit only once to find out which of the two phases was chosen. Can you do that?

2.13.2

Derive the identity

$$(\vec{a}\cdot\vec{\sigma})(\vec{b}\cdot\vec{\sigma}) = (\vec{a}\cdot\vec{b})\mathbf{1} + i(\vec{a}\times\vec{b})\cdot\vec{\sigma}.$$

(All you need here are the Pauli matrices' commutation and anti-commutation relations, but it is instructive to derive the identity using the component notation, and below we give a sketch of how such a derivation would go.)

First, notice that the products of Pauli matrices can be written succinctly as

$$\sigma_i\sigma_j = \delta_{ij}\mathbf{1} + i\varepsilon_{ijk}\,\sigma_k,$$

where δ_{ij} is the Kronecker delta and ε_{ijk} is the **Levi-Civita symbol**:

$$\varepsilon_{ijk} = \begin{cases} +1 & \text{if } (i,j,k) \text{ is } (1,2,3),\ (2,3,1), \text{ or } (3,1,2) \\ -1 & \text{if } (i,j,k) \text{ is } (3,2,1),\ (1,3,2), \text{ or } (2,1,3) \\ \ \ 0 & \text{if } i=j, \text{ or } j=k, \text{ or } k=i \end{cases}$$

That is, ε_{ijk} is 1 if (i,j,k) is an even permutation of $(1,2,3)$, it is -1 if it is an odd permutation, and it is 0 if any index is repeated. The Levi-Civita symbol is anti-symmetric, meaning when any two indices are changed, its sign alternates. Then recall that the scalar (dot) product and vector (cross) product of two Euclidean vectors $\vec{a}$ and $\vec{b}$ can be written, in terms of the components, as

$$\vec{a}\cdot\vec{b} = \sum_{i=1}^{3} a_i b_i$$
$$(\vec{a}\times\vec{b})_i = \sum_{j,k=1}^{3} \varepsilon_{ijk} a_j b_k.$$

The rest is rather straightforward:

$$(\vec{a}\cdot\vec{\sigma})(\vec{b}\cdot\vec{\sigma}) = \sum_{ij} a_i b_j \sigma_i\sigma_j = \ldots.$$

3 Logic and geometry with quantum gates

About understanding the square root of `NOT` *via a physical implementation using* ***symmetric beam-splitters****. More about the Bloch sphere, via the omnipresent* ***Pauli matrices****, which can be described in a more algebraic way.*

Before moving on, we first study two of the subjects from Chapter 2 in more depth: the square root of `NOT`, and the Bloch sphere.

The goal for the latter is to be able to visualise sequences of unitary operations on a qubit as sequences of rotations, and to see the action of some quantum circuits without getting engaged in lengthy calculations. The goal for the former is to study a way of implementing this gate using physical experiments, and then studying a related construction (the so-called **Mach–Zehnder interferometer**).

3.1 Physics against logic, via beamsplitters

A symmetric beam-splitter is a cube of glass which reflects half the light that impinges upon it, while allowing the remaining half to pass through unaffected. For our purposes it can be viewed as a device which has two input and two output ports which we label as $|0\rangle$ and $|1\rangle$, as in Figure 16.

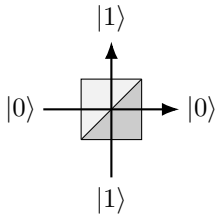


Figure 16. A beam-splitter.

When we aim a single photon at such a beam-splitter using one of the input ports, we notice that the photon doesn't split in two: we can place photo-detectors wherever we like in the apparatus, fire in a photon, and verify that if any of the photo-detectors registers a hit, none of the others do. In particular, if we place a photo-detector behind the beam-splitter in each of the two possible exit beams, the photon is detected with equal probability at either detector, no matter whether the photon was initially fired from input port $|0\rangle$ or $|1\rangle$.

It may seem obvious that, at the very least, the photon is *either* in the transmitted beam $|0\rangle$ *or* in the reflected beam $|1\rangle$ during any one run of this experiment. Thus we may be tempted to think of the beam-splitter as a random binary switch which, with equal probability, transforms any binary input into one of the two possible outputs. However, that is not necessarily the case. Let us introduce a second beam-splitter and place two normal mirrors so that both paths intersect at the second beam-splitter, resulting in something resembling the Mach-Zehnder interferometer (see 17).

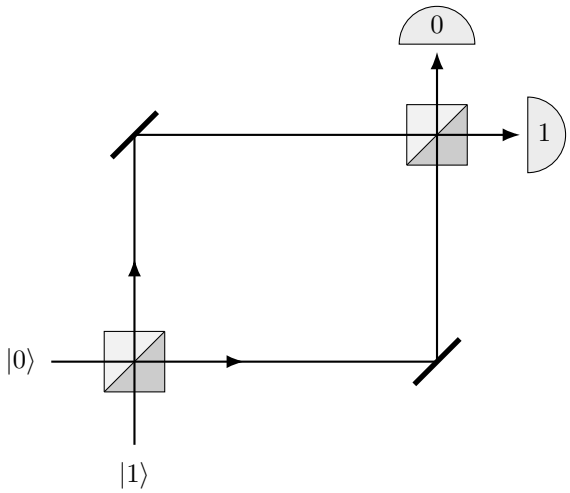
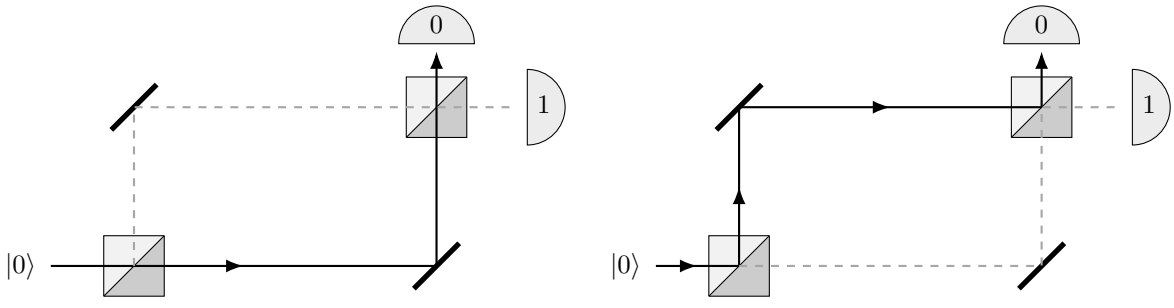


Figure 17. Two beam-splitters with mirrors, so that a photon travels through both.

Now, the axiom of additivity in probability theory says that whenever something can happen in several alternative ways we add probabilities for each way considered separately. We might argue that a photon fired into the input port $|0\rangle$ can reach the detector 0 in two *mutually exclusive* ways: either by two consecutive reflections or by two consecutive transmissions. Each reflection happens with probability $1/2$ and each transmission happens with probability $1/2$ thus the total probability of reaching detector 0 is a sum of the probability of the two consecutive reflections ($1/2 \times 1/2 = 1/4$) and the probability of the two consecutive transmissions ($1/2 \times 1/2 = 1/4$) which gives probability $1/2$. This is summarised in Figure 18, and makes perfect sense — a random switch followed by a random switch should give nothing else but a random switch. However, if we set up such an experiment, that is not what happens!

There is no reason why probability theory (or any other *a priori* mathematical construct for that matter) should make any meaningful statements about outcomes of physical experiments.



two consecutive transmissions:

probability amplitude $= \frac{i}{\sqrt{2}}\frac{i}{\sqrt{2}} = -\frac{1}{2}$

two consecutive reflections:

probability amplitude $= \frac{1}{\sqrt{2}}\frac{1}{\sqrt{2}} = \frac{1}{2}$

Figure 18. The two possible classical scenarios. Note that this is **not** what actually happens in the real physical world!

In experimental reality, when the optical paths between the two beam-splitters are the same, the photon fired from input port $|0\rangle$ *always* strikes detector 1 and *never* detector 0 (and the photon fired from input port $|1\rangle$ *always* strikes detector 0 and *never* detector 1). Thus a beam-splitter acts as the square root of NOT gate.

The action of the beam-splitter — in fact, the action of any quantum device — can be described by tabulating the amplitudes of transitions between its input and output ports.

Note that gate B is not the same square root of NOT as the one described previously. In fact, there are infinitely many ways of implementing this "impossible" logical operation.

$$B = \begin{bmatrix} B_{00} & B_{01} \\ B_{10} & B_{11} \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{i}{\sqrt{2}} \\ \frac{i}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}.$$

The matrix element B_{lk}, where $k, l = 0, 1$, represents the amplitude of transition from input $|k\rangle$ to output $|l\rangle$ (watch the order of indices). Each reflection (entries B_{01} and B_{10}) happens with amplitude $i/\sqrt{2}$ and each transmission (entries B_{00} and B_{11}) happens with amplitude $1/\sqrt{2}$. Thus the total amplitude that a photon fired from input port $|0\rangle$ will reach detector 0 is the sum of the amplitude of the two consecutive reflections ($i/\sqrt{2} \times i/\sqrt{2} = -1/2$) and the amplitude of the two consecutive transmissions ($1/\sqrt{2} \times 1/\sqrt{2} = 1/2$) which gives the total amplitude 0. The resulting probability is then zero. Unlike probabilities, amplitudes can cancel out each other out. We can now go on and calculate the amplitude that the photon will reach detector 1. In this case we will get i, which gives probability 1. We can then switch to input $|1\rangle$ and repeat our calculations. All possible paths and associated amplitudes are shown in 19.

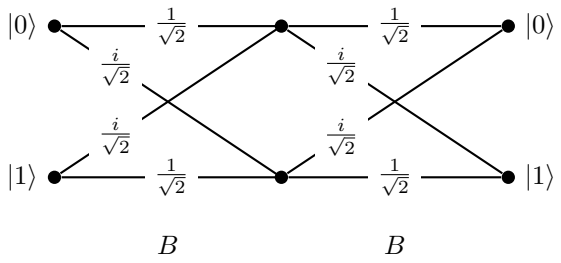


Figure 19. All possible transitions and their amplitudes when we compose two beam-splitters B.

However, instead of going through all the paths in this diagram and linking specific inputs to specific outputs, we can simply multiply the transition matrices:

$$BB = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{i}{\sqrt{2}} \\ \frac{i}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{i}{\sqrt{2}} \\ \frac{i}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} = \begin{bmatrix} 0 & i \\ i & 0 \end{bmatrix} = iX$$

where

$$X = \texttt{NOT} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

Recalling Chapter 2, we see that beam-splitters give a physical way of constructing the square root of NOT.

bit-flip	$\texttt{NOT} \equiv X$	$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$
beam-splitter	$\sqrt{\texttt{NOT}} \equiv B$	$\frac{1}{\sqrt{2}} \begin{bmatrix} 1 & i \\ i & 1 \end{bmatrix}$

3.2 Quantum interference revisited (still about beam-splitters)

One of the simplest quantum devices in which quantum interference can be controlled is a **Mach–Zehnder interferometer** — see Figure 20.

You can play around with a virtual Mach–Zehnder interferometer at Quantum Flytrap. (There are also lots of other things you can do in this virtual lab: go have a look!).

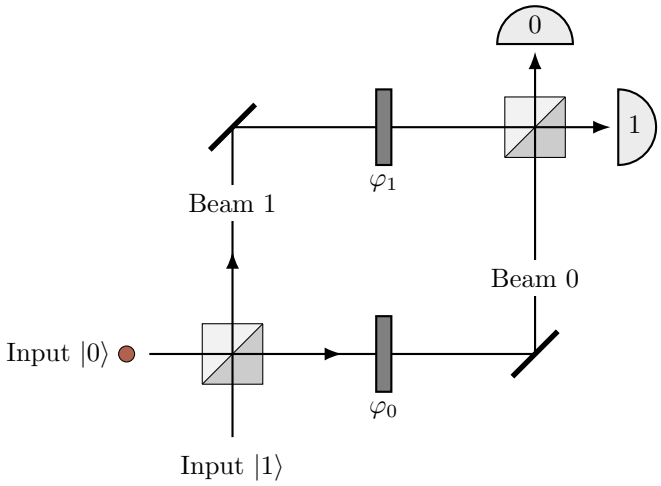


Figure 20. The Mach-Zehnder interferometer, with the input photon represented by the coloured dot. This experimental set-up is named after the physicists Ludwig Mach and Ludwig Zehnder, who proposed it back in 1890s.

It consists of two beam-splitters (the square boxes, bottom left and top right) and two slivers of glass of different thickness which are inserted into each of the optical paths connecting the two beam-splitters. The slivers are usually referred to as "phase shifters" and their thicknesses, φ_0 and φ_1, are measured in units of the photon's wavelength multiplied by 2π. The two inputs ports of the interferometer are labelled as $|0\rangle$ and $|1\rangle$, and each of the two output ports, also labelled as 0 and 1, terminates in a photodetector.

The two diagonal objects in the top-left and bottom-right corners of 20 are simply mirrors to make the two possible paths meet at the second beam-splitter. We will often use i as an index even though it is also used for the imaginary unit. Hopefully, no confusion will arise for it should be clear from the context which one is which.

A photon (the coloured dot in the figure) impinges on the first beam-splitter from one of the two input ports (here input 1) and begins its journey towards one of the two photodetectors. Let U_{ij} denote the probability amplitude that the photon initially in input port $j = 0, 1$ ends up in detector $i = 0, 1$. At each of the two beam-splitters the photon is transmitted with the probability amplitude $\sqrt{T}$ and reflected with the probability amplitude $i\sqrt{R}$, with $R + T = 1$, and the two phase shifters modify the amplitudes by phase factors $e^{i\varphi_0}$ and $e^{i\varphi_1}$, respectively. In quantum theory we almost always start with the amplitudes, and once we have a full expression for the amplitude of a given outcome we square its absolute value to get the corresponding probability.

For example, let us calculate U_{00}. We notice that there are two alternative ways for the photon in the input port 0 to end up in the output port 0. It can take the lower path, through the phase shifter φ_0, or the upper path, through the phase shifter φ_1. The lower

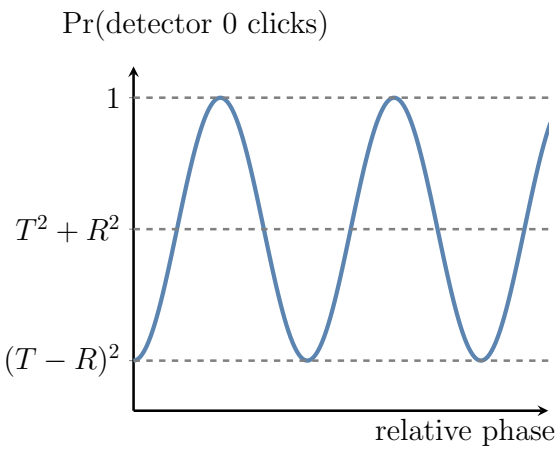
path implies two consecutive transmissions at the beam-splitters and the phase factor $e^{i\varphi_0}$, whereas the upper path implies two consecutive reflections and the phase factor $e^{i\varphi_1}$. Once the photon ends in the output port 0 there is no way of knowing which path was taken, thus we add the amplitudes pertaining to each path. The resulting amplitude is

$$U_{00} = \sqrt{T}e^{i\varphi_0}\sqrt{T} + i\sqrt{R}e^{i\varphi_1}i\sqrt{R},$$

and the corresponding probability $P_{00} = |U_{00}|^2$ reads

$$\begin{aligned} P_{00} &= \left|\sqrt{T}e^{i\varphi_0}\sqrt{T} + i\sqrt{R}e^{i\varphi_1}i\sqrt{R}\right|^2 = \left|Te^{i\varphi_0} - Re^{i\varphi_1}\right|^2 \\ &= T^2 + R^2 - 2TR\cos(\varphi_1 - \varphi_0). \end{aligned}$$

The "classical" part of this expression, $T^2 + R^2$, basically says that the photon undergoes either two consecutive transmissions with probability T^2, or two consecutive reflections with probability R^2. The probability of being transmitted through any phase shifter is always 1, hence the phase shifters play no role in the classical description of this process. But the classical description is not correct, as the experiments show, whence the interference term $2TR\cos(\varphi_1 - \varphi_0)$ in which the phase shifters play the essential role. Depending on the *relative* phase $\varphi = \varphi_1 - \varphi_0$ the probability that the detector 0 "clicks" can vary from $(T - R)^2$, for $\varphi = 0$, to 1, for $\varphi = \pi$.



If we do not care about the experimental details, we can represent the action of the Mach–Zehnder interferometer in terms of a diagram: see 21.

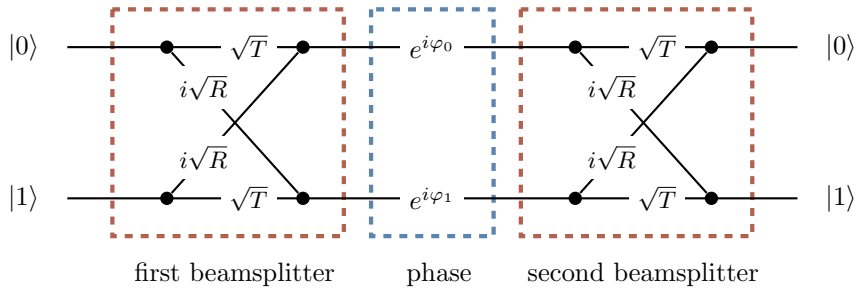


Figure 21. The Mach-Zehnder interferometer represented as an abstract diagram.

Here we can follow, from left to right, the multiple different paths that a photon can take in between specific input and output ports. The amplitude along any given path is just

the product of the amplitudes pertaining to the path segments (Rule 1), while the overall amplitude is the sum of the amplitudes for the many different paths (Rule 2). You can, for example, see that the probability amplitude U_{10} is given by

$$U_{10} = \sqrt{T}e^{i\varphi_0}i\sqrt{R} + i\sqrt{R}e^{i\varphi_1}\sqrt{T}$$

and the corresponding probability is

$$\begin{aligned} P_{10} &= \left|\sqrt{T}e^{i\varphi_0}i\sqrt{R} + i\sqrt{R}e^{i\varphi_1}\sqrt{T}\right|^2 \\ &= 2RT + 2RT\cos(\varphi_1 - \varphi_0). \end{aligned}$$

Again, the first term is of "classical" origin and represents probabilities corresponding to each path: one reflection followed by one transmission plus one transmission followed by one reflection, that is, $RT + TR = 2RT$. The second term is the interference term. Clearly, the photon entering port 1 will end up in one of the two detectors, hence,

$$P_{00} + P_{10} = R^2 + 2RT + T^2 = (T + R)^2 = 1.$$

The action of the interferometer is thus fully described by the four probability amplitudes U_{ij} $(i, j = 0, 1)$.

In general, any isolated quantum device, including a quantum computer, can be described by a matrix of probability amplitudes U_{ij} that input j generates output i. Watch the order of indices.

The most popular instance of a Mach–Zehnder interferometer involves only symmetric beam-splitters ($R = T = \frac{1}{2}$) and is fully described by the matrix

$$U = \begin{bmatrix} -\sin\varphi/2 & \cos\varphi/2 \\ \cos\varphi/2 & \sin\varphi/2 \end{bmatrix}$$

where $\varphi = \varphi_1 - \varphi_0$. In fact, when you do all the calculations you obtain $ie^{i\frac{\varphi_0+\varphi_1}{2}}U$ rather than U, but the global phase factor $ie^{i\frac{\varphi_0+\varphi_1}{2}}$ is common to all the amplitudes in the matrix and as such it does not contribute to the resulting probabilities.

That is, when you write down the matrices describing the action of the symmetric beam-splitters and the phase gates, and then multiply them all together (which is an exercise worth doing).

To better understand why we don't worry about global phase factors, think about the eigenvalues of a matrix A and of the matrix μA, where μ is a complex number with $|\mu| = 1$.

3.3 The Pauli matrices, algebraically

Matrices form a vector space: you can add them, and you can multiply them by a scalar. One possible choice of a basis in the vector space of (2×2) matrices is the set of matrices $\{M_{00}, M_{01}, M_{10}, M_{11}\}$, where the entries of M_{ij} are all 0 except for the ij-th entry, which is 1 (e.g. $M_{01} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$). However, it turns out that there is a different basis which offers lots of insights into the structure of the general single-qubit unitary transformations, namely $\{\mathbf{1}, X, Y, Z\}$, i.e. the identity matrix and the three Pauli matrices. We have already defined the Pauli operators in Chapter 2, but we recall their definition here along with a different notation that we sometimes use.

In this chapter we are concerned only with the *single-qubit* Pauli operators. There are analogous *multi-qubit* Pauli operators, but be careful: these do not satisfy all the same properties! For example, anti-commutativity (explained below) is special to the *single-qubit* case.

Identity	$\mathbf{1} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$
Bit-flip	$X \equiv \sigma_x = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$
Bit-phase-flip	$Y \equiv \sigma_y = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}$
Phase-flip	$Z \equiv \sigma_z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$

Recalling Chapter 2, we know that the Pauli operators (as well as the identity operator) are unitary and Hermitian, square to the identity, and anti-commute.

Any (2×2) complex matrix A has a unique expansion in the form

$$\begin{aligned} A &= \begin{bmatrix} a_0 + a_z & a_x - ia_y \\ a_x + ia_y & a_0 - a_z \end{bmatrix} \\ &= a_0\mathbf{1} + a_x\sigma_x + a_y\sigma_y + a_z\sigma_z \\ &= a_0\mathbf{1} + \vec{a}\cdot\vec{\sigma}. \end{aligned} \tag{3.3.1}$$

$XY + YX = 0,$
$XZ + ZX = 0,$
$YZ + ZY = 0.$

for some complex numbers a_0, a_x, a_y, and a_z. Here, $\vec{a}$ is a vector with three complex components (a_x, a_y, a_z), and $\vec{\sigma}$ represents the "vector" of Pauli matrices $(\sigma_x, \sigma_y, \sigma_z)$. The algebraic properties of the Pauli matrices can be neatly compacted (see the exercises) into a single expression:

The **multiplication rule**:

$$(\vec{a}\cdot\vec{\sigma})\,(\vec{b}\cdot\vec{\sigma}) = (\vec{a}\cdot\vec{b})\,\mathbf{1} + i(\vec{a}\times\vec{b})\cdot\vec{\sigma}.$$

We also introduce the inner product of two matrices:

The **Hilbert–Schmidt product**:

$$(A|B) = \frac{1}{2}\operatorname{tr} A^\dagger B.$$

The $\frac{1}{2}$ coefficient here is simply the normalisation factor, which changes if we consider *multi-qubit* Pauli operators.

Recall that the trace of a square matrix A, denoted by $\operatorname{tr} A$, is defined to be the sum of the elements on the main diagonal of A, and defines a linear mapping: for any scalars α and β,

$$\operatorname{tr}(\alpha A + \beta B) = \alpha \operatorname{tr} A + \beta \operatorname{tr} B.$$

Moreover, the trace is invariant under *cyclic* permutations: e.g.

$$\mathrm{tr}(ABC) = \mathrm{tr}(BCA) = \mathrm{tr}(CAB).$$

Note, however, that this does *not* imply that e.g. $\mathrm{tr}(ABC) = \mathrm{tr}(ACB)$.

3.4 Unitaries as rotations

Now we can finish off what we started in Chapter 2: we know how single-qubit state vectors correspond to points on the Bloch sphere, but now we can study how (2×2) unitary matrices correspond to *rotations* of this sphere.

Geometrically speaking, the group of unitaries $\mathrm{U}(2)$ is a three-dimensional sphere S^3 in $\mathbb{R}^4$. We often fix the determinant to be $+1$ and express (2×2) unitaries as

$$U = u_0\mathbf{1} + i(u_x\sigma_x + u_y\sigma_y + u_z\sigma_z).$$

Such matrices form a popular subgroup of $\mathrm{U}(2)$; it is called the **special** (meaning the determinant is equal to 1) unitary group, and denoted by $\mathrm{SU}(2)$. In quantum theory, any two unitary matrices that differ by some global multiplicative phase factor represent the same physical operation, so we are "allowed to" fix the determinant to be $+1$, and thus restrict ourselves to considering matrices in $\mathrm{SU}(2)$. This is a sensible approach, practised by many theoretical physicists, but again, for some historical reasons, the convention in quantum information science does not follow this approach. For example, phase gates are usually written as

$$P_\alpha = \begin{bmatrix} 1 & 0 \\ 0 & e^{i\alpha} \end{bmatrix}$$

rather than

$$P_\alpha = \begin{bmatrix} e^{-i\frac{\alpha}{2}} & 0 \\ 0 & e^{i\frac{\alpha}{2}} \end{bmatrix}$$

Still, sometimes the T gate

$$T = \begin{bmatrix} 1 & 0 \\ 0 & e^{i\pi/4} \end{bmatrix} = \begin{bmatrix} e^{-i\pi/8} & 0 \\ 0 & e^{i\pi/8} \end{bmatrix}$$

is called the $\pi/8$ gate, because of its $\mathrm{SU}(2)$ form.

So let us write any (2×2) unitary, up to an overall phase factor, as

$$U = u_0\mathbf{1} + i(u_x\sigma_x + u_y\sigma_y + u_z\sigma_z) = u_0\mathbf{1} + i\vec{u}\cdot\vec{\sigma}$$

where $u_0^2 + |\vec{u}|^2 = 1$. This additional unitarity restriction allows us to parametrise u_0 and $\vec{u}$ in terms of a real unit vector $\vec{n}$, parallel to $\vec{u}$, and a real angle θ so that

As you can see, we often make progress and gain insights simply by choosing a convenient parametrisation.

$$U = (\cos\theta)\mathbf{1} + (i\sin\theta)\vec{n}\cdot\vec{\sigma}.$$

An alternative way of writing this expression is

$$U = e^{i\theta\vec{n}\cdot\vec{\sigma}},$$

Although this looks neat and tidy, the normal convention is to parametrise so that $U = e^{i\frac{-\theta}{2}\vec{n}\cdot\vec{\sigma}}$, and then the direction follows from the right-hand rule, and the rotation corresponds to that in the Bloch sphere. It is a useful exercise to show that you can write U in this way as well.

as follows from the power-series expansion of the exponential. Indeed, any unitary matrix can always be written in the exponential form as

$$e^{iA} \equiv \mathbf{1} + iA + \frac{(iA)^2}{1\cdot 2} + \frac{(iA)^3}{1\cdot 2\cdot 3}\cdots$$
$$= \sum_{n=0}^{\infty} \frac{(iA)^n}{n!}$$

where A is a Hermitian matrix.

Writing unitary matrices in the form e^{iA} is analogous to writing complex numbers of unit modulus as $e^{i\alpha}$ (the so-called **polar form**).

Now comes a remarkable connection between two-dimensional unitary matrices and ordinary three-dimensional rotations:

The unitary $U = e^{i\theta\vec{n}\cdot\vec{\sigma}}$ represents a clockwise rotation through the angle 2θ about the axis defined by $\vec{n}$. (**N.B.** 2θ, *not* θ).

For example,

$$e^{i\theta\sigma_x} = \begin{bmatrix} \cos\theta & i\sin\theta \\ i\sin\theta & \cos\theta \end{bmatrix}$$
$$e^{i\theta\sigma_y} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}$$
$$e^{i\theta\sigma_z} = \begin{bmatrix} e^{i\theta} & 0 \\ 0 & e^{-i\theta} \end{bmatrix}$$

represent rotations by 2θ about the x-, y- and z-axis, respectively.

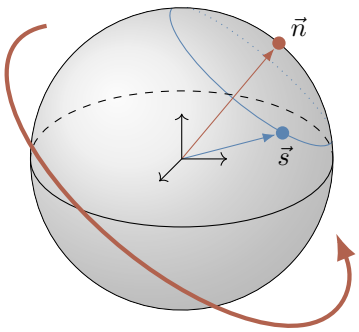


Figure 22. $e^{i\theta\vec{n}\cdot\vec{\sigma}}$ rotates the vector $\vec{s}$ about $\vec{n}$ by angle 2θ, sending it to a point on the blue circle, whose centre is passed through by $\vec{n}$.

Now we can show that the Hadamard gate

$$\begin{aligned} H &= \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \\ &= \frac{1}{\sqrt{2}}(\sigma_x + \sigma_z) \\ &= (-i)e^{i\frac{\pi}{2}\frac{1}{\sqrt{2}}(\sigma_x+\sigma_z)} \end{aligned}$$

(which, up to the overall multiplicative phase factor of $-i$, is equal to $e^{i\frac{\pi}{2}\frac{1}{\sqrt{2}}(\sigma_x+\sigma_z)}$) represents rotation about the diagonal $(x+z)$-axis through the angle π.

In somewhat abstract terms, we make the connection between unitaries and rotations by looking how the unitary group $\mathrm{U}(2)$ acts on the three-dimensional Euclidian space of (2×2) Hermitian matrices *with zero trace*. All such matrices S can be written as $S = \vec{s}\cdot\vec{\sigma}$ for some real $\vec{s}$, i.e. each matrix is represented by a Euclidean vector $\vec{s}$ in $\mathbb{R}^3$. Now, $U \in \mathrm{U}(2)$ acts on the Euclidean space of such matrices by $S \mapsto S' = USU^\dagger$, i.e.

$$\vec{s}\cdot\vec{\sigma} \mapsto \vec{s'}\cdot\vec{\sigma} = U(\vec{s}\cdot\vec{\sigma})U^\dagger \tag{3.5.1}$$

This is a linear map $\mathbb{R}^3 \to \mathbb{R}^3$, and is thus given by some (3×3) real-valued matrix:

$$R_U : \mathbb{R}^3 \to \mathbb{R}^3. \tag{3.5.2}$$

We note that this map is an isometry (a distance preserving operation), since it preserves the scalar product in the Euclidean space: for any two vectors $\vec{s}$ and $\vec{v}$,

$$\begin{aligned} \vec{s'}\cdot\vec{v'} &= \frac{1}{2}\operatorname{tr}[S'V'] \\ &= \frac{1}{2}\operatorname{tr}[(USU^\dagger)(UVU^\dagger)] \\ &= \frac{1}{2}\operatorname{tr}[SV] \\ &= \vec{s}\cdot\vec{v} \end{aligned}$$

(where $S = \vec{s}\cdot\vec{\sigma}$ and $V = \vec{v}\cdot\vec{\sigma}$) using the cyclic property of the trace. This means that the matrix R_U is *orthogonal*.

Orthogonal transformations preserve the length of vectors as well as the angles between them.

Furthermore, we can show that $\det R_U = 1$. But the only isometries in three dimensional Euclidian space (which are described by orthogonal matrices with determinant 1) are rotations.

We can also see that $\det R_U = 1$ from the fact that any matrix in $\mathrm{U}(2)$ can be smoothly connected to the identity.

Thus, in the mathematical lingo, we have established a homomorphism

$$\begin{aligned} \mathrm{U}(2) &\longrightarrow \mathrm{SO}(3) \\ U &\longmapsto R_U \end{aligned}$$

Recall that a homomorphism is a structure-preserving map between two algebraic structures of the same type, in our case two groups. An isomorphism between algebraic structures of the same type is one-to-one homomorphism.

where $\mathrm{SO}(3)$ stands for the special orthogonal group in three dimensions (the group of all rotations about the origin of three-dimensional Euclidean space $\mathbb{R}^3$ under the operation of composition). It is quite clear from Equation (3.5.1) that unitary matrices differing only by a global multiplicative phase factor (e.g. U and $e^{i\gamma}U$) represent the same rotation.

Physicists, however, usually prefer a more direct demonstration, which goes roughly like this. Consider the map $\vec{s} \mapsto \vec{s'}$ induced by $U = e^{i\alpha\vec{n}\cdot\vec{\sigma}}$. For small values of α, we can

write

$$\begin{aligned}\vec{s'}\cdot\vec{\sigma} &= U(\vec{s}\cdot\vec{\sigma})U^\dagger \\ &= \Big(\mathbf{1}+i\alpha(\vec{n}\cdot\vec{\sigma})+\dots\Big)(\vec{s}\cdot\vec{\sigma})\Big(\mathbf{1}-i\alpha(\vec{n}\cdot\vec{\sigma})+\dots\Big).\end{aligned}$$

To the first order in α, this gives

$$\vec{s'}\cdot\vec{\sigma} = \Big(\vec{s}+2\alpha(\vec{n}\times\vec{s})\Big)\cdot\vec{\sigma}$$

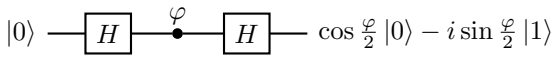
that is,

$$\vec{s'} = \vec{s}+2\alpha(\vec{n}\times\vec{s})$$

which we recognise as a good old textbook formula for an infinitesimal clockwise rotation of $\vec{s}$ about the axis $\vec{n}$ through the angle 2α.

3.5 Universality, again

Although this may all seem tediously abstract, it is surprisingly useful. Take another look at the single qubit interference circuit

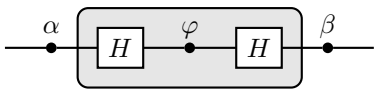


and the corresponding sequence of unitary operations

$$\begin{aligned}H\left(e^{-i\frac{\varphi}{2}Z}\right)H &= e^{-i\frac{\varphi}{2}X} \\ &= \begin{bmatrix}\cos\varphi/2 & -i\sin\varphi/2\\ -i\sin\varphi/2 & \cos\varphi/2\end{bmatrix}\end{aligned}$$

The single qubit interference circuit has a simple geometrical meaning: it shows how a rotation about the z-axis, induced by the phase gate P_φ, is turned, by the two Hadamard gates, into a rotation about the x-axis.

Now, take a look at this circuit:



What does it represent? The central part is a rotation by φ about the x-axis, but is it is sandwiched between two rotations about the z-axis. Now we have to appeal to your knowledge of classical mechanics: you may recall that any rotation in the Euclidean space can be performed as a sequence of three rotations: one about z-axis, one about x-axis, and one more about the z-axis. In this context, this implies that any unitary U, up to a global phase factor, can be written as

$$\begin{aligned}U(\alpha,\beta,\varphi) &= e^{-i\frac{\beta}{2}Z}e^{-i\frac{\varphi}{2}X}e^{-i\frac{\alpha}{2}Z} \\ &= \begin{bmatrix}e^{-i\left(\frac{\alpha+\beta}{2}\right)}\cos\frac{\varphi}{2} & ie^{i\left(\frac{\alpha-\beta}{2}\right)}\sin\frac{\varphi}{2}\\ ie^{-i\left(\frac{\alpha-\beta}{2}\right)}\sin\frac{\varphi}{2} & e^{i\left(\frac{\alpha+\beta}{2}\right)}\cos\frac{\varphi}{2}\end{bmatrix}.\end{aligned}$$

Thus once you are given a couple of Hadamard gates and an infinite supply of phase gates, so that you can choose the three phases you need, you can construct an arbitrary unitary operation on a single qubit. Needless to say, the two axes in question, z and x, do not have any special status, geometrically speaking — if we have rotations about any two orthogonal axes then we can create any one-qubit unitary that we want.

In fact, even this condition isn't necessary! See Figure 23

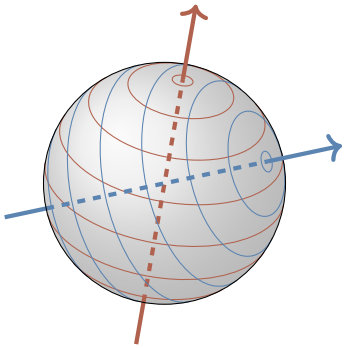
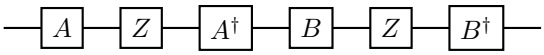


Figure 23. If we can move along the two families of circles, then from any point on the sphere we can reach any other point. The two axes do not even have to be orthogonal: any two different axes will do. Can you see why?

Consider the following circuit:



where both A and B are unitary operations. We claim that *any* unitary U can be represented in this form.

Again, we can prove this geometrically. The circuit represents two rotations by $180°$ about two axes obtained by rotating the z-axis with unitaries A and B, respectively. Any rotation in the three-dimensional space is the composition of two rotations by $180°$, as shown in Figure 24. The resulting axis of rotation is perpendicular to the two axes about which rotations by $180°$ are performed, and the angle of the composed rotation is twice the angle between the two axes.

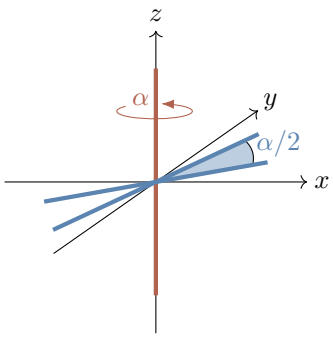


Figure 24. Rotating by α around the z-axis is the same as the composition of two rotations by $180°$ around axes which both lie in the xy-plane, with angle $\alpha/2$ between them.

3.6 Some quantum dynamics

The time evolution of a quantum state is a unitary process which is generated by a Hermitian operator called the **Hamiltonian**, which we also denote by H. The Hamiltonian

Hopefully it will always be clear from the context which H refers to Hamiltonian and which H to Hadamard. Don't confuse the two!

contains a complete specification of all interactions within the system under consideration. In an isolated system, the state vector $|\psi(t)\rangle$ changes smoothly in time according to the **Schrödinger equation**:

$$\frac{\mathrm{d}}{\mathrm{d}t}|\psi(t)\rangle = -\frac{i}{\hbar}H|\psi(t)\rangle.$$

For **time-independent** Hamiltonians (i.e. where $|\psi(t)\rangle = |\psi\rangle$ has no t-dependence), the formal solution of this reads

$$\begin{aligned} &|\psi(t)\rangle = U(t)|\psi(0)\rangle \\ &\text{where} \quad U(t) = e^{-\frac{i}{\hbar}Ht}. \end{aligned}$$

Here $\hbar$ denotes **Planck's constant**, which has the value $\hbar = 1.05 \times 10^{-34}$ Js. However, theorists always choose to work with a system of units where $\hbar = 1$.

Now, to relate this to the earlier parts of this chapter, we note that the Hamiltonian of a qubit can always be written in the form $H = E_0\mathbf{1} + \omega(\vec{n}\cdot\vec{\sigma})$, hence

$$\begin{aligned} U(t) &= e^{-i\omega t\vec{n}\cdot\vec{\sigma}} \\ &= (\cos\omega t)\mathbf{1} - (i\sin\omega t)\vec{n}\cdot\vec{\sigma} \end{aligned}$$

which is a rotation with angular frequency ω about the axis defined by the unit vector $\vec{n}$.

!!!TODO!!! the 4π world of qubits

3.7 Remarks and Exercises

3.7.1

Show that $\{\mathbf{1}, \sigma_x, \sigma_y, \sigma_z\}$ is an orthonormal basis with respect to the Hilbert-Schmidt product in the space of complex (2×2) matrices.

3.7.2

Show that the coefficients a_k (for $k = x, y, z$) in Equation (3.3.1) are given by the inner product $a_k = (\sigma_k|A) = \frac{1}{2}\operatorname{tr}\sigma_k A$.

3.7.3

1. Show that $\frac{1}{2}\operatorname{tr}(\vec{a}\cdot\vec{\sigma})(\vec{b}\cdot\vec{\sigma}) = \vec{a}\cdot\vec{b}$.
2. Show that any $\vec{n}\cdot\vec{\sigma}$ has eigenvalues $\pm|\vec{n}|$.
3. Show that, if $\vec{n}\cdot\vec{m} = 0$, then the operators $\vec{n}\cdot\vec{\sigma}$ and $\vec{m}\cdot\vec{\sigma}$ anticommute.

3.7.4

In these notes, we usually deal with matrices that are Hermitian ($A = A^\dagger$) or unitary ($AA^\dagger = \mathbf{1}$). It is easy to see that, if A is Hermitian, then a_0 and the three components of $\vec{a}$ are all real. The (2×2) unitaries are usually parametrised as

$$U = e^{i\gamma}\Big(u_0\mathbf{1} + i(u_x\sigma_x + u_y\sigma_y + u_z\sigma_z)\Big)$$

where $e^{i\gamma}$ is an overall multiplicative phase factor, with γ real, and u_0 and the three components u_x, u_y, u_z are all real numbers.

Show that the unitarity condition implies that

$$u_0^2 + u_x^2 + u_y^2 + u_z^2 = 1,$$

and show that the determinant of U is $e^{i2\gamma}$ using this parametrisation.

3.7.5

1. Show that, if $A^2 = \mathbf{1}$, then we can manipulate the power series expansion of e^{iA} into a simple expression: for any real α,

$$e^{i\alpha A} = (\cos\alpha)\mathbf{1} + (i\sin\alpha)A.$$

2. Show that any (2×2) unitary matrix U can be written, up to an overall multiplicative phase factor, as

$$U = e^{i\theta\vec{n}\cdot\vec{\sigma}} = (\cos\theta)\mathbf{1} + (i\sin\theta)\vec{n}\cdot\vec{\sigma}.$$

(The argument here is the same as the argument that $e^{i\theta} = \cos\theta + i\sin\theta$).

3.7.6

Show that $\operatorname{tr}\sigma_x\sigma_y\sigma_z = 2i$.

3.7.7

1. Consider

$$U(\vec{e}_k\cdot\sigma_k)U^\dagger = U\sigma_k U^\dagger = \vec{f}_k\cdot\vec{\sigma}.$$

So U maps the unit vectors $\vec{e}_x$, $\vec{e}_y$, and $\vec{z}_z$, (along the x-, y-, and z-axis, respectively), to new unit vectors $\vec{f}_x$, $\vec{f}_y$, and $\vec{f}_z$. We already know that, in Euclidean space, this transformation is described by a (3×3) orthogonal matrix R_U. How are the three vectors $\vec{f}_x$, $\vec{f}_y$, and $\vec{f}_z$ related to the entries in matrix R_U?

2. Show that

$$\begin{aligned}\operatorname{tr}\sigma_x\sigma_y\sigma_z &= \operatorname{tr}(\vec{f}_x\cdot\vec{\sigma})(\vec{f}_y\cdot\vec{\sigma})(\vec{f}_z\cdot\vec{\sigma})\\ &= 2i\det R_U\end{aligned}$$

(which implies that $\det R_U = 1$).

3. Make use of the orthonormality of the Pauli basis and Equation (3.5.1) to show that the elements of the matrix $R = R_U$ can be expressed in terms of those of the matrix U, in the form

$$R_{ij} = \frac{1}{2}\operatorname{tr}\left(\sigma_i U\sigma_j U^\dagger\right).$$

Here, i and j take values in $\{1,2,3\}$, and $\sigma_1\equiv\sigma_x$, $\sigma_2\equiv\sigma_y$, $\sigma_3\equiv\sigma_z$.

3.7.8

Show that the phase gate

$$P_\varphi = \begin{bmatrix}1 & 0\\ 0 & e^{i\varphi}\end{bmatrix}$$

represents an anticlockwise rotation about the z-axis through the angle φ.

Hint. It might be helpful to start with the $\mathrm{SU}(2)$ version of the phase gate:

$$\begin{aligned} P_\varphi &= e^{-i\frac{\varphi}{2}\sigma_z} \\ &= \begin{bmatrix} e^{-i\frac{\varphi}{2}} & 0 \\ 0 & e^{i\frac{\varphi}{2}} \end{bmatrix} \quad \longrightarrow \quad R \\ &= \begin{bmatrix} \cos\varphi & -\sin\varphi & 0 \\ \sin\varphi & \cos\varphi & 0 \\ 0 & 0 & 1 \end{bmatrix} \end{aligned}$$

3.7.9

1. Express the Hadamard gate H in terms of $\vec{n}\cdot\vec{\sigma}$, and show that

$$\begin{aligned} HZH &= X \\ HXH &= Z \\ HYH &= -Y. \end{aligned}$$

2. Show that the Hadamard gate H turns rotations about the x-axis into rotations about the z-axis, and vice versa. That is,

$$\begin{aligned} H\left(e^{-i\frac{\varphi}{2}Z}\right)H &= e^{-i\frac{\varphi}{2}X} \\ H\left(e^{-i\frac{\varphi}{2}X}\right)H &= e^{-i\frac{\varphi}{2}Z}. \end{aligned}$$

3.7.10 Swiss Granite Fountain

In the Singapore Botanic Gardens, there is a sculpture by Ueli Fausch called "Swiss Granite Fountain". It is a spherical granite ball which measures 80cm in diameter and weighs 700kg, and is kept afloat by strong water pressure directed through the basal block. It is easy to set the ball in motion, and it keeps rotating in whatever way you start for a long time. Suppose you are given access to this ball only near the top, so that you can push it to make it rotate around any horizontal axis, but you don't have enough of a grip to make it turn around the vertical axis. Can you make it rotate around the vertical axis anyway?

3.7.11

A qubit (spin one-half particle) initially in state $|0\rangle$ (spin up) is placed in a uniform magnetic field. The interaction between the field and the qubit is described by the Hamiltonian

$$H = \omega \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}$$

where ω is proportional to the strength of the field. What is the state of the qubit after time $t = \pi/4\omega$?

In Earth's magnetic field, which is about 0.5 gauss, the value of ω is of the order of 10^6 cycles per second.

4 Measurements

*About the **Hilbert-space formalism** of quantum theory, and the role of **measurements** in quantum information theory, as well as introducing the quantum dramas of Alice and Bob.*

Eventually, we have to talk about quantum measurements, since, at some point, someone has to look at a measuring device and register the outcome of whatever quantum circuits we've been designing. It turns out that this is a bit more tricky than one might think. Quantum measurement is not a passive acquisition of information: if you measure, you disturb. Even though it is a physical process, like any other quantum evolution, it is traditionally described by a different set of mathematical tools.

4.1 Hilbert spaces, briefly

A formal mathematical setting for a quantum system is that of a **Hilbert space** $\mathcal{H}$, i.e. a vector space with an inner product. The result of any preparation of a system is represented by some unit vector $|\psi\rangle \in \mathcal{H}$, and any test is represented by some other unit vector $|e\rangle \in \mathcal{H}$. The inner product of these two vectors, $\langle e|\psi\rangle$, gives the probability amplitude that an object prepared in state $|\psi\rangle$ will pass a test for being in state $|e\rangle$. Probabilities are obtained by squaring absolute values of probability amplitudes:

$$|\langle e|\psi\rangle|^2 = \langle\psi|e\rangle\langle e|\psi\rangle.$$

The term "Hilbert space" used to be reserved for an infinite-dimensional inner product space that is **complete**, i.e. such that every Cauchy sequence in the space converges to an element in the space. Nowadays, as in these notes, the term includes finite-dimensional spaces, which automatically satisfy the condition of completeness.

After the test, in which the object was found to be in state $|e\rangle$, say, the object forgets about its previous state $|\psi\rangle$ and is, indeed, actually now in state $|e\rangle$. This is the mysterious **quantum collapse** which we will briefly discuss later on.

A more complete test involves multiple states e_k that form an orthonormal basis $\{|e_1\rangle, \ldots, |e_n\rangle\}$. These states are perfectly distinguishable from each other: the condition $\langle e_k|e_l\rangle = \delta_{kl}$ implies that a quantum system prepared in state $|e_l\rangle$ will never be found in state $|e_k\rangle$ (unless $k = l$). The probability amplitude that the system in state $|\psi\rangle$ will be found in state $|e_k\rangle$ is $\langle e_k|\psi\rangle$ and, given that the vectors $|e_k\rangle$ span the whole vector space, the system will be always found in one of the basis states, whence

$$\sum_k |\langle e_k|\psi\rangle|^2 = 1.$$

As a result:

A **complete** measurement in quantum theory is determined by the choice of an orthonormal basis $\{|e_i\rangle\}$ in $\mathcal{H}$, and every such basis (in principle) represents a possible complete measurement.

4.2 Back to qubits; complete measurements

A **projector** is any Hermitian operator $P = P^\dagger$ which is idempotent ($P^2 = P$). The rank of P is evaluated using $\mathrm{tr}(P)$. In the Dirac notation, $|e\rangle\langle e|$ is a rank one projector on the subspace spanned by the unit vector $|e\rangle$, and it acts on any vector $|v\rangle$ as $(|e\rangle\langle e|)|v\rangle = |e\rangle\langle e|v\rangle$.

The most common measurement in quantum information science is the standard measurement on a qubit, also referred to as the measurement in the **standard** (or **computational**) basis $\{|0\rangle, |1\rangle\}$. When we draw circuit diagrams it is tacitly assumed that such a measurement is performed on each qubit at the end of quantum evolution.

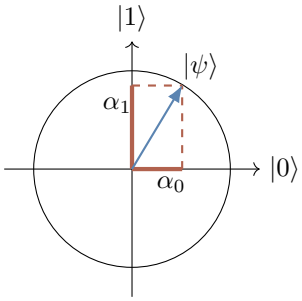


Figure 25. The standard (computational) basis defines the standard measurements.

However, if we want to emphasise the role of the measurement, then we can include it explicitly in the diagram as a special quantum gate, e.g. as

$$|\psi\rangle = \alpha_0\,|0\rangle + \alpha_1\,|1\rangle \longrightarrow \begin{cases} |0\rangle & \text{with probability } |\alpha_0|^2 \\ |1\rangle & \text{with probability } |\alpha_1|^2 \end{cases}$$

or, in an alternative notation, as

$$|\psi\rangle = \alpha_0\,|0\rangle + \alpha_1\,|1\rangle \longrightarrow \boxed{k} \longrightarrow |k\rangle \quad \text{with probability } |\alpha_k|^2 \quad (k = 0, 1)$$

As we can see, if the qubit is prepared in state $|\psi\rangle = \alpha_0|0\rangle + \alpha_1|1\rangle$ and subsequently measured in the standard basis state, then the outcome is $|k\rangle$ (for $k = 0, 1$) with probability

This slick argument is a good example of how nice the bra-ket notation can be.

$$\begin{aligned} |\alpha_k|^2 &= |\langle k|\psi\rangle|^2 \\ &= \underbrace{\langle\psi|k\rangle}_{\alpha_k^\star}\underbrace{\langle k|\psi\rangle}_{\alpha_k} \\ &= \langle\psi|\underbrace{|k\rangle\langle k|}_{\text{projector}}|\psi\rangle \\ &= \langle\psi|P_k|\psi\rangle \end{aligned}$$

where $P_k = |k\rangle\langle k|$ is the projector on $|k\rangle$. If the outcome of the measurement is k, then the output state of the measurement gate is $|k\rangle$. The original state $|\psi\rangle$ is *irretrievably lost*. This sudden change of the state, from the pre-measurement state $|\psi\rangle$ to the post-measurement state, either $|0\rangle$ or $|1\rangle$, is often called a **collapse** or a **reduction** of the state.

So it looks like there are two distinct ways for a quantum state to change: on the one hand we have unitary evolutions, and on the other hand we have an abrupt change during the measurement process. Surely, the measurement process is not governed by any different laws of physics?

No, it is not!

A measurement is a physical process and can be explained without any "collapse", but it is usually a complicated process in which one complex system (a measuring apparatus or an observer) interacts and gets correlated with a physical system being measured. We will discuss this more later on, but for now let us accept a "collapse" as a *convenient mathematical shortcut*, and describe it in terms of projectors rather than unitary operators.

4.3 The projection rule; incomplete measurements

So far we have identified measurements with orthonormal bases, or, if you wish, with a set of orthonormal projectors on the basis vectors.

- The **orthonormality** condition:

  $$\langle e_k|e_l\rangle = \delta_{kl}$$

  i.e. the basis consists of unit vectors that are pairwise orthogonal.
- The **completeness** condition:

  $$\sum_k |e_k\rangle\langle e_k| = \mathbf{1}$$

  i.e. *any* vector in $\mathcal{H}$ can be expressed as the sum of the orthogonal projections on the $|e_k\rangle$.

Given a quantum system in state $|\psi\rangle$ such that $|\psi\rangle = \sum_k \alpha_k|e_k\rangle$, we can write

$$\begin{aligned}|\psi\rangle &= \mathbf{1}|\psi\rangle \\ &= \sum_k (|e_k\rangle\langle e_k|)|\psi\rangle \\ &= \sum_k |e_k\rangle\langle e_k|\psi\rangle \\ &= \sum_k |e_k\rangle\alpha_k \\ &= \sum_k \alpha_k|e_k\rangle.\end{aligned}$$

This says that the measurement in the basis $\{|e_i\rangle\}$ gives the outcome labelled by e_k with probability

$$|\langle e_k|\psi\rangle|^2 = \langle\psi|e_k\rangle\langle e_k|\psi\rangle$$

and leaves the system in state $|e_k\rangle$. This is a *complete* measurement, which represents the best we can do in terms of resolving state vectors in the basis states. But sometimes we do not want our measurement to distinguish all the elements of an orthonormal basis.

For example, a complete measurement in a four-dimensional Hilbert space will have four distinct outcomes: $|e_1\rangle$, $|e_2\rangle$, $|e_3\rangle$, and $|e_4\rangle$, but we may want to lump together some of the outcomes and distinguish, say, only between $\{|e_1\rangle, |e_2\rangle\}$, and $\{|e_3\rangle, |e_4\rangle\}$. In other words, we might be trying to distinguish one *subspace* from another, without separating

vectors that lie in the same subspace. Such measurements (said to be **incomplete**) are indeed possible, and they can be less disruptive than the complete measurements.

Intuitively, an incomplete measurement has fewer outcomes and is hence less informative, but the state after such a measurement is usually less disturbed.

In general, instead of projecting on one dimensional subspaces spanned by vectors from an orthonormal basis, we can decompose our Hilbert space into mutually orthogonal subspaces of various dimensions and project on them.

- The orthogonality conditions for projectors:

$$P_k P_l = P_k \delta_{kl}$$

- The projector decomposition of the identity:

$$\sum_k P_k = \mathbf{1}$$

For any decomposition of the identity into orthogonal projectors P_k, there exists a measurement that takes a quantum system in state $|\psi\rangle$, outputs label k with probability $\langle\psi|P_k|\psi\rangle$, and leaves the system in the state $P_k|\psi\rangle$ (multiplied by the normalisation factor, i.e. divided by the length of $P_k|\psi\rangle$):

$$|\psi\rangle \mapsto \frac{P_k|\psi\rangle}{\sqrt{\langle\psi|P_k|\psi\rangle}}.$$

4.4 Example of an incomplete measurement

Consider a three-dimensional Hilbert space $\mathcal{H}$, and the two orthogonal projectors

$$P = |e_1\rangle\langle e_1| + |e_2\rangle\langle e_2|$$
$$Q = |e_3\rangle\langle e_3|$$

that form the decomposition of the identity: $P + Q = \mathbf{1}$. Suppose that a physical system is prepared in state $|\psi\rangle = \alpha_1|e_1\rangle + \alpha_2|e_2\rangle + \alpha_3|e_3\rangle$. Ideally, we would like to perform a complete measurement that would resolve the state $|v\rangle$ into the three basis states, but suppose our experimental apparatus is not good enough, and lumps together $|e_1\rangle$ and $|e_2\rangle$. In other words, it can only differentiate between the two subspaces associated with projectors P and Q.

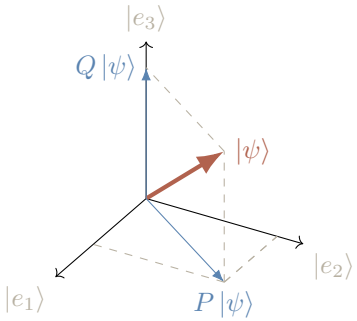
The apparatus, in this incomplete measurement, may find the system in the subspace associated with P. This happens with probability

$$\begin{aligned}\langle\psi|P|\psi\rangle &= \langle\psi|e_1\rangle\langle e_1|\psi\rangle + \langle\psi|e_2\rangle\langle e_2|\psi\rangle \\ &= |\alpha_1|^2 + |\alpha_2|^2,\end{aligned}$$

and the state right after the measurement is the normalised vector $P|\psi\rangle$, i.e.

$$\frac{\alpha_1|e_1\rangle + \alpha_2|e_2\rangle}{\sqrt{|\alpha_1|^2 + |\alpha_2|^2}}.$$

The measurement may also find the system in the subspace associated with Q with the probability $\langle\psi|Q|\psi\rangle = |\alpha_3|^2$, resulting in the post-measurement state $|e_3\rangle$.



4.5 Observables

An **observable** A is a measurable physical property which has a numerical value, for example, spin or momentum or energy. The term "observable" also extends to any basic measurement in which each outcome is associated with a numerical value. If λ_k is the numerical value associated with outcome $|e_k\rangle$ then we say that the observable A is **represented** by the operator

$$\begin{aligned} A &= \sum_k \lambda_k |e_k\rangle\langle e_k| \\ &= \sum_k \lambda_k P_k, \end{aligned}$$

where λ_k is now the eigenvalue corresponding to the eigenvector $|e_k\rangle$, or the projector P_k.

Recall the following types of operator:

normal	$AA^\dagger = A^\dagger A$
unitary	$AA^\dagger = A^\dagger A = \mathbf{1}$
Hermitian, or **self-adjoint**	$A^\dagger = A$
positive semi-definite	$\langle v\|A\|v\rangle \geqslant 0$ for all $\|v\rangle$

Note that an operator A is normal if and only if it is unitarily diagonalisable, and that both unitary and Hermitian operators are normal.

Conversely, to any normal operator A we can associate a measurement defined by the eigenvectors of A, which form an orthonormal basis, and use the eigenvalues of A to label the outcomes of this measurement. If we choose the eigenvalues to be real numbers then A becomes a Hermitian operator. For example, the standard measurement on a single qubit is often called the Z**-measurement**, because the Pauli Z operator can be diagonalised in the standard basis and written as $Z = (+1)|0\rangle\langle 0| + (-1)|1\rangle\langle 1|$. The two outcomes, $|0\rangle$ and

$|1\rangle$, are now labelled as $+1$ and -1, respectively. Using the same association we also have the X- and the Y-measurements, defined by the Pauli X and Y operators, respectively.

> The outcomes can be labelled by any symbols of your choice; it is the *decomposition* of the Hilbert space into *mutually orthogonal subspaces* that defines a measurement, not the labels.

This said, labelling outcomes with real numbers is very useful. For example, the expected value $\langle A\rangle$, which is the average of the numerical values λ_k weighted by their probabilities, is a very useful quantity and can be easily expressed in terms of the operator A as $\langle\psi|A|\psi\rangle$, as follows:

Some textbooks describe observables in terms of Hermitian operators, claiming that the corresponding operators *have to* be Hermitian "because the outcomes are real numbers". This is actually a bit backwards. As we say above, the labels can be arbitrary, but, since real number labels are often useful (as we're about to justify), we tend to only work with Hermitian operators.

$$\begin{aligned}\sum_k \lambda_k \Pr(\text{outcome k}) &= \sum_k \lambda_k |\langle e_k|\psi\rangle|^2 \\ &= \sum_k \lambda_k \langle\psi|e_k\rangle\langle e_k|\psi\rangle \\ &= \langle\psi| \left(\sum_k \lambda_k |e_k\rangle\langle e_k|\right) |\psi\rangle \\ &= \langle\psi|A|\psi\rangle.\end{aligned}$$

To be clear, this is not a value we expect to see in any particular experiment. Instead, imagine a huge number of quantum objects, all prepared in the state $|\psi\rangle$ and think about the observable A being measured on each of the objects. Statistically we then expect the average of our measurement results to be roughly $\langle A\rangle$. Note that when A is a projector then $\langle\psi|A|\psi\rangle$ is the probability of the outcome associated with A.

4.6 Compatible observables and the uncertainty relation

!!!TODO!!!

4.7 Quantum communication

Now is a good moment to introduce Alice and Bob (not their real names): our two protagonists who always need to communicate with each other. These two play the major role in many communication dramas, though they remain rather lacking in character development.

This time Alice is sending quantum states to Bob, and Bob does his best to identify them correctly by choosing appropriate measurements. Let us start we a simple observation: if a quantum state of the carrier of information is described by a state vector in a 2^n-dimensional Hilbert space, then the carrier can carry at most n bits of information. For example, Alice can choose one of the 2^n states from a pre-agreed orthonormal basis $\{|e_k\rangle\}$, and Bob will be able to distinguish them reliably by choosing the $\{|e_k\rangle\}$ basis for his measurement.

But can Alice and Bob do better than that? Can Alice send more than n bits of information per carrier by encoding them in states $|s_1\rangle, \ldots, |s_N\rangle$ where $N \geqslant 2^n$? Can Bob choose a clever measurement and reliably distinguish between all such states?

The answer is *no*.

4.8 Basic quantum coding and decoding

Suppose Alice randomly chooses one of the pre-agreed N signal states $|s_k\rangle$ and sends it to Bob, who tries to identify the signal states by performing a measurement defined by the projectors P_l. Let P be a projector on a subspace spanned by the signal states $|s_k\rangle$, i.e. $P|s_k\rangle = |s_k\rangle$. The dimension d of this subspace is given by $d = \operatorname{tr} P$. We shall assume, without any loss of generality, that Bob designed his measurement in such a way that, whenever he gets outcome P_k, he concludes that Alice sent state $|s_k\rangle$. His probability of success is given by

$$\Pr(\text{success}) = \frac{1}{N}\sum_k \langle s_k|P_k|s_k\rangle$$

which is the probability that signal state $|s_k\rangle$ is selected (here equal to $1/N$, since all the signal states are equally likely) times the probability that the selected signal state is correctly identified by Bob (which is $\langle s_k|P_k|s_k\rangle$), and we sum over all signal states.

We have the following **trace identities**:

- $\operatorname{tr}(ABC) = \operatorname{tr}(BCA) = \operatorname{tr}(CAB)$
- $\operatorname{tr}|a\rangle\langle b| = \langle b|a\rangle$
- $\operatorname{tr}|a\rangle\langle b| = \langle b|A|a\rangle$
- $\operatorname{tr} BP \leqslant \operatorname{tr} B$ for any positive semi-definite B and projector P.

To prove this last identity, consider the projector $Q = \mathbf{1} - P$, and note that

$$\begin{aligned}\operatorname{tr} B &= \operatorname{tr} B(P+Q)\\ &= \operatorname{tr} BP + \operatorname{tr} BQ\end{aligned}$$

and that $\operatorname{tr} BQ$ is non-negative.

Let us use this case to practice some of the trace identities. It is often convenient to write expressions such as $\langle\psi|A|\psi\rangle$ in terms of the trace: for any vector $|\psi\rangle$ and operator A we have

$$\begin{aligned}\langle\psi|A|\psi\rangle &= \operatorname{tr}(A|\psi\rangle\langle\psi|)\\ &= \operatorname{tr}(|\psi\rangle\langle\psi|A).\end{aligned}$$

In our case,

$$\begin{aligned}\Pr(\text{success}) &= \frac{1}{N}\sum_k \langle s_k|P_k|s_k\rangle\\ &= \frac{1}{N}\sum_k \langle s_k|PP_kP|s_k\rangle\\ &= \frac{1}{N}\sum_k \operatorname{tr}(PP_kP|s_k\rangle\langle s_k|)\end{aligned}$$

where we have also used that $P|s_k\rangle = |s_k\rangle$. Let us bound this expression above by using

the aforementioned trace identities:

$$\begin{aligned}\sum_k \frac{1}{N}\langle s_k|P_k|s_k\rangle &= \frac{1}{N}\sum_k \operatorname{tr}(PP_kP|s_k\rangle\langle s_k|)\\ &\leqslant \frac{1}{N}\sum_k \operatorname{tr}(PP_kP)\\ &= \frac{1}{N}P\left(\sum_k P_k\right)P\\ &= \frac{1}{N}\operatorname{tr}(P\mathbf{1}P)\\ &= \frac{1}{N}\operatorname{tr}(P)\\ &= \frac{d}{N}.\end{aligned}$$

So if Alice encodes N equally likely messages as states in a quantum system that, mathematically speaking, lives in the Hilbert space of dimension d, and if Bob decodes by performing a measurement and inferring the message from the result, then Bob's probability of success is bounded by $\frac{d}{N}$. If the number N of possible signals exceeds the dimension d, then Bob will not be able to reliably distinguish between the signals by any measurement. In particular:

> In this setting, one qubit can store *at most* one bit of information that can *reliably* be read by a measurement.

There is something called **superdense coding**, where one qubit can actually store *two* classical bits, but this relies on Alice and Bob both having access to a shared entangled state right from the very start of the experiment.

4.9 Distinguishability of non-orthogonal states

We have already mentioned that non-orthogonal states cannot be reliably distinguished, and now it is time to make this statement more precise. Suppose Alice sends Bob a message by choosing one of the two non-orthogonal states $|s_1\rangle$ and $|s_2\rangle$, where both are equally likely to be chosen. What is the probability that Bob will decode the message correctly and what is the best (i.e. the one that maximises this probability) measurement? As a general rule, before you embark on any calculations, check for symmetries, special cases, and anything that may help you to visualise the problem and make intelligent guesses about the solution. One of the most powerful research tools is a good guess. In fact, this is what real research is about: educated guesses that guide your calculations. In this particular case you can use symmetry arguments to guess the optimal measurement — see Figure 26. Once you have guessed the answer, you might as well do the calculations.

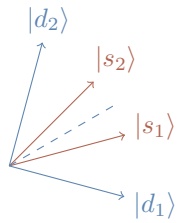


Figure 26. The optimal measurement to distinguish between the two equally likely non-orthogonal signal states $|s_1\rangle$ and $|s_2\rangle$ is described by the two orthogonal vectors $|d_1\rangle$ and $|d_2\rangle$ placed symmetrically around the signal states.

Suppose Bob's measurement is described by projectors P_1 and P_2, with the inference rule "P_1 implies $|s_1\rangle$; P_2 implies $|s_2\rangle$". Then

$$\begin{aligned}\Pr(\text{success}) &= \frac{1}{2}\left(\langle s_1|P_1|s_1\rangle + \langle s_2|P_2|s_2\rangle\right)\\ &= \frac{1}{2}\left(\operatorname{tr} P_1|s_1\rangle\langle s_1| + \operatorname{tr} P_2|s_2\rangle\langle s_2|\right)\\ &= \frac{1}{2}\left(\operatorname{tr} P_1|s_1\rangle\langle s_1| + \operatorname{tr}(\mathbf{1}-P_1)|s_2\rangle\langle s_2|\right)\\ &= \frac{1}{2}\left(1+\operatorname{tr} P_1\left(|s_1\rangle\langle s_1| - |s_2\rangle\langle s_2|\right)\right).\end{aligned}$$

Let us look at the operator $D = |s_1\rangle\langle s_1| - |s_2\rangle\langle s_2|$ that appears in the last expression. This operator acts on the subspace spanned by $|s_1\rangle$ and $|s_2\rangle$; it is Hermitian; the sum of its two (real) eigenvalues is zero; and $\operatorname{tr} D = \langle s_1|s_1\rangle - \langle s_2|s_2\rangle = 0$. Let us write D as $\lambda(|d_+\rangle\langle d_+| - |d_-\rangle\langle d_-|)$, where $|d_\pm\rangle$ are the two orthonormal eigenstates of D, and $\pm\lambda$ are the corresponding eigenvalues. Now we write

$$\begin{aligned}\Pr(\text{success}) &= \frac{1}{2}\left(1+\lambda\operatorname{tr} P_1\left(|d_+\rangle\langle d_+| - |d_-\rangle\langle d_-|\right)\right)\\ &\leqslant \frac{1}{2}\left(1+\lambda\langle d_+|P_1|d_+\rangle\right)\end{aligned}$$

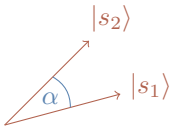
where we have dropped the non-negative term $\operatorname{tr} P_1|d_-\rangle\langle d_-|$. In fact, it is easy to see that we will maximise the expression above by choosing $P_1 = |d_+\rangle\langle d_+|$ (and $P_2 = |d_-\rangle\langle d_-|$). The probability of success is then bounded by $\frac{1}{2}(1+\lambda)$. All we have to do now is to find the positive eigenvalue λ for the operator D. We can do this, of course, by solving the characteristic equation for a matrix representation of D, but, as we are now practising the trace operations, we can also notice that $\operatorname{tr} D^2 = 2\lambda^2$, and then evaluate the trace of D^2. We use the trace identities and obtain

$$\begin{aligned}\operatorname{tr} D^2 &= \operatorname{tr}\left(|s_1\rangle\langle s_1| - |s_2\rangle\langle s_2|\right)\left(|s_1\rangle\langle s_1| - |s_2\rangle\langle s_2|\right)\\ &= 2 - 2|\langle s_1|s_2\rangle|^2\end{aligned}$$

which gives $\lambda = \sqrt{1-|\langle s_1|s_2\rangle|^2}$. Bringing it all together we have the final expression:

$$\Pr(\text{success}) = \frac{1}{2}\left(1+\sqrt{1-|\langle s_1|s_2\rangle|^2}\right).$$

We can parametrise $|\langle s_1|s_2\rangle| = \cos\alpha$, and interpret α as the angle between $|s_1\rangle$ and $|s_2\rangle$.



This allows us to express our findings in a clearer way: given two equally likely states, $|s_1\rangle$ and $|s_2\rangle$, such that $|\langle s_1|s_2\rangle| = \cos\alpha$, the probability of correctly identifying the state by a projective measurement is bounded by

$$\Pr(\text{success}) = \frac{1}{2}(1+\sin\alpha),$$

and the optimal measurement that achieves this bound is determined by the eigenvectors of $D = |s_1\rangle\langle s_1| - |s_2\rangle\langle s_2|$ (try to visualise these eigenvectors).

It makes sense, right? If we try just guessing the state, without any measurement, then we expect $\Pr(\text{success}) = \frac{1}{2}$. This is our lower bound, and in any attempt to distinguish the two states we should do better than that. If the two signal states are very close to each other then $\sin\alpha$ is small and we are slightly better off than guessing. As we increase α, the two states become more distinguishable, and, as we can see from the formula, when the two states become orthogonal they also become completely distinguishable.

4.10 Wiesner's quantum money

!!!TODO!!!

4.11 Remarks and Exercises

4.11.1

Consider two unit vectors $|a\rangle$ and $|b\rangle$. Is $|a\rangle\langle a| + |b\rangle\langle b|$ a projector?

4.11.2

Suppose you are given a single qubit in some unknown quantum state $|\psi\rangle$. Can you determine $|\psi\rangle$?

4.11.3

You measure a random qubit in the standard basis and register $|0\rangle$. What does it tell you about the pre-measurement state $|\psi\rangle$?

4.11.4

How many real parameters do you need to determine $|\psi\rangle$? Would you be able to reconstruct $|\psi\rangle$ from $\langle\psi|X|\psi\rangle$, $\langle\psi|Y|\psi\rangle$, and $\langle\psi|Z|\psi\rangle$? (It may help you to visualise $|\psi\rangle$ as a Bloch vector).

4.11.5

You are given zillions of qubits, all prepared in the same quantum state $|\psi\rangle$. How would you determine $|\psi\rangle$?

4.11.6

The Z measurement is defined by the projectors

$$P_0 = \frac{1}{2}(\mathbf{1} + Z),$$
$$P_1 = \frac{1}{2}(\mathbf{1} - Z).$$

Consider the measurement associated to some Hermitian operator S that satisfies $S^2 = \mathbf{1}$. Show that the two outcomes ± 1 correspond to the projectors $\frac{1}{2}(\mathbf{1} \pm S)$.

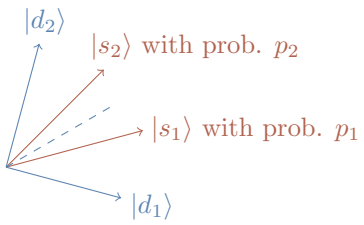
4.11.7

In our quantum circuits, unless specified otherwise, all measurements are assumed to be performed in the standard basis. This is because any measurement can be reduced to the standard measurement by performing some prior unitary transformation. Show that any two orthonormal bases $\{|e_k\rangle\}$ and $\{|d_l\rangle\}$ are always related by some unitary U (i.e. show that $\sum_k |d_k\rangle\langle e_k|$ is unitary).

Suppose projectors P_k define the standard measurement, and show that for any unitary U projectors $UP_kU^\dagger$ also define a measurement.

4.11.8

The optimal measurement to distinguish between the two equally likely non-orthogonal signal states, $|s_1\rangle$ and $|s_2\rangle$, is described by the two orthogonal vectors $|d_1\rangle$ and $|d_2\rangle$, placed symmetrically around the signal states. But suppose the states are not equally likely: say $|s_1\rangle$ is chosen with probability p_1 and $|s_2\rangle$ with probability p_2. How would you modify the measurement to maximise the probability of success in this case?



4.11.9 The values of σ_x and σ_y of a qubit

(This is a simplified version of a beautiful quantum puzzle proposed in 1987 by Lev Vaidman, Yakir Aharonov, and David Z. Albert in a paper with the somewhat provocative title "How to ascertain the values of σ_x, σ_y, and σ_z of a spin-$\frac{1}{2}$ particle". For the original, see *Phys. Rev. Lett.* **58** (1987), 1385.)

Alice prepares a qubit in any state of her choosing and gives it to Bob, who secretly measures either σ_x or σ_y. The outcome of the measurement is seen only by Bob. Alice has no clue which measurement was chosen by Bob, but right after his measurement she gets her qubit back and she can measure it as well. Some time later, Bob tells Alice which of the two measurements was chosen, i.e. whether he measured σ_x or σ_y. Alice then tells him the outcome he obtained in his measurement. Bob is surprised, since the two measurements have mutually unbiased bases and yet Alice always gets it right. How does she do it?

4.11.10 The Zeno effect

!!!TODO!!!

4.12 Quantum theory, formally

Even though multiplying and adding probability amplitudes is essentially all there is to quantum theory, we hardly ever multiply and add amplitudes in a pedestrian way. Instead, as we have seen, we neatly tabulate the amplitudes into vectors and matrices and let the matrix multiplication take care of multiplication and addition of amplitudes corresponding to different alternatives. Thus vectors and matrices appear naturally as our bookkeeping tools: we use vectors to describe quantum states, and matrices (operators) to describe quantum evolutions and measurements. This leads to a convenient mathematical setting for quantum theory, which is a complex vector space with an inner product, often referred to as a Hilbert space. It turns out, somewhat miraculously, that this pure mathematical construct is exactly what we need to formalise quantum theory. It gives us a precise language which is appropriate for making empirically testable predictions. At a very instrumental level, quantum theory is a set of rules designed to answer questions such as "given a specific preparation and a subsequent evolution, how can we compute probabilities for the

outcomes of such-and-such measurement". Here is how we represent preparations, evolutions and measurements in mathematical terms, and how we get probabilities.

Note that we have already said much of the below, but we are summarising it again now in a more precise way, formally defining the mathematical framework of quantum theory that we use.

We also need to point out that a vital part of the formalism of quantum theory is missing from the following description, namely the idea of **tensor products**. To talk about this, we need to introduce the notion of **entanglement**, and this will be the subject of Chapter 5.

4.12.1 Quantum states

With any isolated quantum system which can be prepared in n perfectly distinguishable states, we can associate a Hilbert space $\mathcal{H}$ of dimension n such that each vector $|v\rangle \in \mathcal{H}$ of unit length ($\langle v|v\rangle = 1$) represents a quantum state of the system. The overall phase of the vector has no physical significance: $|v\rangle$ and $e^{i\varphi}|v\rangle$, for any real φ, describe the same state. The inner product $\langle u|v\rangle$ is the probability amplitude that a quantum system prepared in state $|v\rangle$ will be found in state $|u\rangle$. States corresponding to orthogonal vectors, $\langle u|v\rangle = 0$, are perfectly distinguishable, since the system prepared in state $|v\rangle$ will never be found in state $|u\rangle$, and vice versa. In particular, states forming orthonormal bases are always perfectly distinguishable from each other.

4.12.2 Quantum evolutions

> Any physically admissible evolution of an isolated quantum system is represented by a unitary operator.

Unitary operators describing evolutions of quantum systems are usually derived from the **Schrödinger equation**.

Recall our previous discussion of this equation in Chapter 3.

$$\frac{\mathrm{d}}{\mathrm{d}t}|\psi(t)\rangle = -\frac{i}{\hbar}H|\psi(t)\rangle$$

where H is a Hermitian operator called the Hamiltonian.

This equation contains a complete specification of all interactions both within the system and between the system and the external potentials. For time independent Hamiltonians, the formal solution of the Schrödinger equation reads

$$\begin{aligned} &|\psi(t)\rangle = U(t)|\psi(0)\rangle \\ &\text{where} \quad U(t) = e^{-\frac{i}{\hbar}Ht}. \end{aligned}$$

Any unitary matrix can be represented as the exponential of some Hermitian matrix H and some real coefficient t:

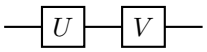
We ignore convergence issues.

$$\begin{aligned} e^{-itH} &\equiv \mathbf{1} - itH + \frac{(-it)^2}{2}H^2 + \frac{(-it)^3}{2\cdot 3}H^3 + \dots \\ &= \sum_{n=0}^{\infty} \frac{(-it)^n}{n!}H^n. \end{aligned}$$

The state vector changes smoothly: for $t = 0$ the time evolution operator is merely the unit operator $\mathbf{1}$, and when t is very small $U(t) \approx \mathbf{1} - itH$ is close to the unit operator, differing from it by something of order t.

4.12.3 Quantum circuits

In this course we will hardly refer to the Schrödinger equation. Instead we will assume that our clever colleagues, experimental physicists, are able to implement certain unitary operations and we will use these unitaries, like lego blocks, to construct other, more complex, unitaries. We refer to preselected elementary quantum operations as **quantum logic gates** and we often draw diagrams, called **quantum circuits**, to illustrate how they act on qubits. For example, two unitaries, U followed by V, acting on a single qubit are represented as



This diagram should be read from left to right, and the horizontal line represents a qubit that is inertly carried from one quantum operation to another.

4.12.4 Measurements

A complete measurement in quantum theory is determined by the choice of an orthonormal basis $\{|e_i\rangle\}$ in $\mathcal{H}$, and every such basis (in principle) represents a possible measurement. Given a quantum system in state $|\psi\rangle$ such that

$$|\psi\rangle = \sum_i |e_i\rangle\langle e_i|\psi\rangle,$$

the measurement in the basis $\{|e_i\rangle\}$ gives the outcome labelled by e_k with probability $|\langle e_k\psi||\rangle^2$, and leaves the system in state $|e_k\rangle$. This is consistent with our interpretation of the inner product $\langle e_k|\psi\rangle$ as the probability amplitude that a quantum system prepared in state $|\psi\rangle$ will be found in state $|e_k\rangle$. State vectors forming orthonormal bases are perfectly distinguishable from each other ($\langle e_i|e_j\rangle = \delta_{ij}$), so there is no ambiguity about the outcome. A complete measurement is the best we can do in terms of resolving state vectors in the basis states.

In general, for any decomposition of the identity $\sum_k P_k = \mathbf{1}$ into orthogonal projectors P_k (i.e. $P_kP_l = P_k\delta_{kl}$), there exists a measurement that takes a quantum system in state $|\psi\rangle$, outputs label k with probability $\langle\psi|P_k|\psi\rangle$, and leaves the system in the state $P_k|\psi\rangle$ (multiplied by the normalisation factor i.e. divided by the length of $P_k|\psi\rangle$):

$$|\psi\rangle \mapsto \frac{P_k|\psi\rangle}{\sqrt{\langle\psi|P_k|\psi\rangle}}.$$

The projector formalism covers both complete and incomplete measurements. The complete measurements are defined by rank one projectors, $P_k = |e_k\rangle\langle e_k|$, projecting on vectors from some orthonormal basis $\{|e_k\rangle\}$.

5 Quantum entanglement

About the fundamental tool of quantum computing: ***entanglement****, via the formalism of* ***tensor products****, which was the missing ingredient from our previous formalism of quantum theory. Also about various* ***controlled gates****, including the always useful* ***controlled-NOT****.*

We now know everything we need to know about a single qubit and its quantum behaviour. But if we want to understand quantum computation — a complicated quantum interference of many interacting qubits — then we will need few more mathematical tools. Stepping up from one qubit to two or more is a bigger leap than you might expect. Already, with just two qubits, we will encounter the remarkable phenomenon of quantum entanglement and have a chance to discuss some of the most puzzling features of quantum theory that took people decades to understand.

5.1 A small history

The notion of **quantum entanglement** was the subject of many early debates that focused on the meaning of quantum theory. Back in the 1930s, Albert Einstein, Niels Bohr, Werner Heisenberg, and Erwin Schrödinger (to mention just the usual suspects) were trying hard to understand its conceptual consequences. Einstein, the most sceptical of them all, claimed that it was pointing toward the fatal flaw in quantum theory, and referred to it as "*spooky action-at-a-distance*". In contrast, Schrödinger was much more prepared to accept quantum theory exactly as it was formulated, along with all its predictions, no matter how weird they might be. In his 1935 paper, which introduced quantum entanglement, he wrote "I would not call it *one* but rather *the* characteristic trait of quantum mechanics, the one that enforces its entire departure from classical lines of thought". Today we still talk a lot about quantum entanglement, but more often it is viewed as a physical resource which enables us to communicate with perfect security, build very precise atomic clocks, and even teleport small quantum objects! But what exactly is quantum entanglement?

E. Schrödinger, Discussion of probability relations between separated system, *Mathematical Proceedings of the Cambridge Philosophical Society* **31** (1935), pp. 555–563.

5.2 One, two, many...

In classical physics, the transition from a single object to a composite system of many objects is trivial: in order to describe the state of, say, 42 objects at any given moment of time, it is sufficient to describe the state of each of the objects separately. Indeed, the classical state of 42 point-like particles is described by specifying the position and the momentum of each particle.

> In the *classical* world, "the whole is the sum of its parts", but the *quantum* world is very different.

Consider, for example, a pair of qubits. Suppose that each one is described by a state vector: the first one by $|a\rangle$, and the second one by $|b\rangle$. One might therefore think that the most general state of the two qubits should be represented by a pair of state vectors, $|a\rangle|b\rangle$, with one for each qubit. Indeed, such a state is certainly possible, but there are other states that *cannot* be expressed in this form. In order to write down the most general state of two qubits we first focus on the basis states.

For a single qubit we have been using the standard basis $\{|0\rangle, |1\rangle\}$. For two qubits we may choose the following as our standard basis states:

$$|00\rangle \equiv |0\rangle|0\rangle \;\; |01\rangle \equiv |0\rangle|1\rangle$$
$$|10\rangle \equiv |1\rangle|0\rangle \;\; |11\rangle \equiv |1\rangle|1\rangle.$$

It looks like we are defining some sort of "multiplication rule" for kets here, saying that $|a\rangle|b\rangle := |ab\rangle$. This is indeed the case, but to talk about this properly we need to introduce the idea of **tensor products** (which we do very soon).

Within each ket, the first symbol refers to the first qubit, and the second to the second, and we have tacitly assumed that we can distinguish the two qubits by their location, or some other means. Now, the most general state of the two qubits is a normalised linear combination of these four basis states, i.e. a vector of the form

$$|\psi\rangle = c_{00}|00\rangle + c_{01}|01\rangle + c_{10}|10\rangle + c_{11}|11\rangle.$$

Physical interpretation aside, let us count how many real parameters are needed to specify this state. Six, right? We have four complex numbers (eight real parameters) restricted by the normalisation condition, along with the fact that states differing only by a global phase factor are equivalent, which leaves us with six real parameters. Now, by the same line of argument, we need only two real parameters to specify the state of a single qubit, and hence need four real parameters to specify any state of two qubits of the form $|a\rangle|b\rangle$. So it cannot be the case that every state of two qubits can be expressed as a pair of states $|a\rangle|b\rangle$, simply for "dimension reasons".

For example, compare the two states of two qubits,

$$\frac{1}{\sqrt{2}}|00\rangle + \frac{1}{\sqrt{2}}|01\rangle$$
and
$$\frac{1}{\sqrt{2}}|00\rangle + \frac{1}{\sqrt{2}}|11\rangle.$$

The first one is **separable**, i.e. we can view it as a pair of state vectors where each one pertains to one of the two qubits:

$$\frac{1}{\sqrt{2}}|00\rangle + \frac{1}{\sqrt{2}}|01\rangle = \underbrace{|0\rangle}_{\text{qubit 1}} \; \underbrace{\frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)}_{\text{qubit 2}},$$

The second state, however, does *not* admit such a decomposition: there do *not* exist any ψ_1, ψ_2 such that

$$\frac{1}{\sqrt{2}}|00\rangle + \frac{1}{\sqrt{2}}|11\rangle = |\psi_1\rangle|\psi_2\rangle$$

and so we say that it is an **entangled** state. Informally, any bipartite state that cannot be viewed as a pair of two states pertaining to the constituent subsystems is said to be entangled.

With this discussion in our minds, we can now give more formal account of the states of composite quantum systems.

5.3 Quantum theory, formally (continued)

Recalling Chapter 4, where we said that we were missing a key part of the formalism of quantum theory, we can now fill in this hole. Our mathematical formalism of choice behind the quantum theory of composite systems is based on the **tensor product** of Hilbert spaces.

5.3.1 Tensor products

Let the states of some system $\mathcal{A}$ be described by vectors in an n-dimensional Hilbert space $\mathcal{H}_\mathcal{A}$, and the states of some system $\mathcal{B}$ by vectors in an m-dimensional Hilbert space $\mathcal{H}_\mathcal{B}$. The combined system of $\mathcal{A}$ and $\mathcal{B}$ is then described by vectors in the (nm)-dimensional **tensor product space** $\mathcal{H}_\mathcal{A} \otimes \mathcal{H}_\mathcal{B}$. Given bases $\{|a_1\rangle, \ldots, |a_n\rangle\}$ of $\mathcal{H}_\mathcal{A}$ and $\{|b_1\rangle, \ldots, |b_m\rangle\}$ of $\mathcal{H}_\mathcal{B}$, we form a basis of the tensor product by taking the ordered pairs $|a_i\rangle \otimes |b_j\rangle$, for $i = 1, \ldots, n$ and $j = 1, \ldots, m$. The tensor product space $\mathcal{H}_\mathcal{A} \otimes \mathcal{H}_\mathcal{B}$ then consists of all linear combination of such tensor product basis vectors:

If the bases $\{|a_i\rangle\}$ and $\{|b_j\rangle\}$ are orthonormal then so too is the tensor product basis $\{|a_i\rangle \otimes |b_j\rangle\}$.

$$|\psi\rangle = \sum_{ij} c_{ij} |a_i\rangle \otimes |b_j\rangle. \tag{5.3.1}$$

The tensor product operation $\otimes$ is distributive:

$$|a\rangle \otimes (\beta_1|b_1\rangle + \beta_2|b_2\rangle) = \beta_1|a\rangle \otimes |b_1\rangle + \beta_2|a\rangle \otimes |b_2\rangle$$
$$(\alpha_1|a_1\rangle + \alpha_2|a_2\rangle) \otimes |b\rangle = \alpha_1|a_1\rangle \otimes |b\rangle + \alpha_2|a_2\rangle \otimes |b\rangle.$$

The tensor product of Hilbert spaces is again a Hilbert space: the inner products on $\mathcal{H}_\mathcal{A}$ and $\mathcal{H}_\mathcal{B}$ give a natural inner product on $\mathcal{H}_\mathcal{A} \otimes \mathcal{H}_\mathcal{B}$, defined for any two product vectors by

$$(\langle a'| \otimes \langle b'|)\,(|a\rangle \otimes |b\rangle) = \langle a'|a\rangle\langle b'|b\rangle$$

and extended by linearity to sums of tensor products of vectors, and, by associativity, to any number of subsystems. Note that the bra corresponding to the tensor product state $|a\rangle \otimes |b\rangle$ is written as $(|a\rangle \otimes |b\rangle)^\dagger = \langle a| \otimes \langle b|$, where the order of the factors on either side of $\otimes$ does not change when the dagger operation is applied.

$(\mathcal{H}_a \otimes \mathcal{H}_b) \otimes \mathcal{H}_c = \mathcal{H}_a \otimes (\mathcal{H}_b \otimes \mathcal{H}_c)$.

Some joint states of $\mathcal{A}$ and $\mathcal{B}$ can be expressed as a single tensor product, say $|\psi\rangle = |a\rangle \otimes |b\rangle$ (often written as $|a\rangle|b\rangle$, or $|a, b\rangle$, or even $|ab\rangle$), meaning that the subsystem $\mathcal{A}$ is in state $|a\rangle$, and the subsystem $\mathcal{B}$ in state $|b\rangle$. If we expand $|a\rangle = \sum_i \alpha_i|a_i\rangle$ and $|b\rangle = \sum_i \beta_j|b_j\rangle$, then $|\psi\rangle = \sum_{ij} \alpha_i\beta_j|a_i\rangle \otimes |b_j\rangle$ and we see that, for all such states, the coefficients c_{ij} in Equation (5.3.1) are of a rather special form:

$$c_{ij} = \alpha_i\beta_j.$$

We call such states **separable** (or just **product states**). States that are not separable are said to be **entangled**.

A useful fact about tensor products is that $\lambda a \otimes b = a \otimes \lambda b$ (where a and b are vectors, and λ is a scalar). This means that we don't need to worry about brackets, and can write something like $\lambda(a \otimes b)$.

We will also need the concept of the tensor product of two operators. If A is an operator on $\mathcal{H}_\mathcal{A}$ and B an operator on $\mathcal{H}_\mathcal{B}$, then the tensor product operator $A \otimes B$ is an operator on $\mathcal{H}_\mathcal{A} \otimes \mathcal{H}_\mathcal{B}$ defined by its action on product vectors via

$$(A \otimes B)(|a\rangle \otimes |b\rangle) = (A|a\rangle) \otimes (B|b\rangle)$$

and with its action on all other vectors determined by linearity:

$$A \otimes B \left(\sum_{ij} c_{ij} |a_i\rangle \otimes |b_j\rangle \right) = \sum_{ij} c_{ij} A|a_i\rangle \otimes B|b_j\rangle.$$

5.4 Back to qubits

Let's see how this formalism works for qubits. The n-fold tensor product of vectors from the standard basis $\{|0\rangle, |1\rangle\}$ represent binary strings of length n. For example, for $n = 3$,

$$|0\rangle \otimes |1\rangle \otimes |1\rangle \equiv |011\rangle$$
$$|1\rangle \otimes |1\rangle \otimes |1\rangle \equiv |111\rangle.$$

We often drop the $\otimes$ symbol, especially when we deal with the standard tensor product basis. For example, a state of a quantum register composed of four qubits holding the binary string 1001 may be written as $|1\rangle \otimes |0\rangle \otimes |0\rangle \otimes |1\rangle$, as $|1\rangle|0\rangle|0\rangle|1\rangle$, or even simply as $|1001\rangle$.

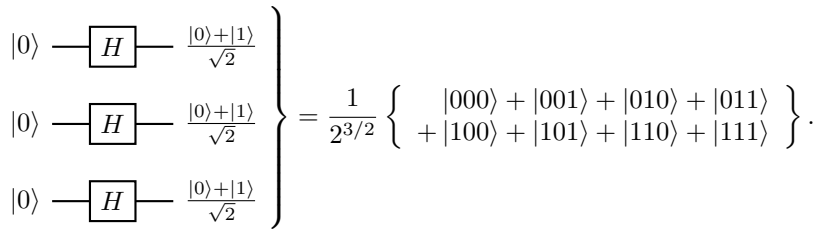
A *classical* register composed of three bits can store *only one* of these two binary strings at any time; a *quantum* register composed of three qubits can store *both of them* in a superposition.

Indeed, if we start with the state $|011\rangle$ and apply the Hadamard gate to the first qubit (which is the same as applying $H \otimes \mathbf{1} \otimes \mathbf{1}$), then, given that linear combinations distribute over tensor products, we obtain

We often simply write $H \otimes H \otimes H$ as $H^{\otimes 3}$.

$$\begin{aligned} |011\rangle &\longmapsto \frac{1}{\sqrt{2}}\big(|0\rangle + |1\rangle\big) \otimes |1\rangle \otimes |1\rangle \\ &\equiv \frac{1}{\sqrt{2}}\big(|011\rangle + |111\rangle\big). \end{aligned}$$

In fact, we can even prepare this register in a superposition of all eight possible binary strings: if we apply the tensor product operation $H \otimes H \otimes H$ to the state $|0\rangle \otimes |0\rangle \otimes |0\rangle = |000\rangle$ then we get



$$= \frac{1}{2^{3/2}} \left\{ \begin{array}{r} |000\rangle + |001\rangle + |010\rangle + |011\rangle \\ + |100\rangle + |101\rangle + |110\rangle + |111\rangle \end{array} \right\}.$$

The resulting state is exactly a superposition of all binary string of length 3, and can also be written as

$$\frac{1}{\sqrt{2}}\big(|0\rangle + |1\rangle\big) \otimes \frac{1}{\sqrt{2}}\big(|0\rangle + |1\rangle\big) \otimes \frac{1}{\sqrt{2}}\big(|0\rangle + |1\rangle\big).$$

In general, the tensor product operation $H^{\otimes n}$, which means "apply the Hadamard gate to each of your n qubits", is known as the **Hadamard transform**, and it maps product states to product states. Like the Hadamard gate in the typical quantum interference circuit, the Hadamard transform opens and closes a multi-qubit interference.

5.5 Separable or entangled?

Most vectors in $\mathcal{H}_a \otimes \mathcal{H}_b$ are entangled and *cannot* be written as product states $|a\rangle \otimes |b\rangle$ with $|a\rangle \in \mathcal{H}_a$ and $|b\rangle \in \mathcal{H}_b$.

In order to see this, let us write any joint state $|\psi\rangle$ of $\mathcal{A}$ and $\mathcal{B}$ in a product basis as

$$\begin{aligned} |\psi\rangle &= \sum_{ij} c_{ij}|a_i\rangle \otimes |b_j\rangle \\ &= \sum_i |a_i\rangle \otimes \left(\sum_j c_{ij}|b_j\rangle\right) \\ &= \sum_i |a_i\rangle \otimes |\phi_i\rangle \end{aligned} \tag{5.5.1}$$

where the $|\phi_i\rangle = \sum_j c_{ij}|b_j\rangle$ are vectors in $\mathcal{H}_\mathcal{B}$ that need not be normalised. For any *product* state, these vectors have a special form. Indeed, if $|\psi\rangle = |a\rangle \otimes |b\rangle$ then, after expanding the first state in the $|a_i\rangle$ basis, we obtain

$$|\psi\rangle = \sum_i |a_i\rangle \otimes \left(\sum_i \alpha_i|b\rangle\right).$$

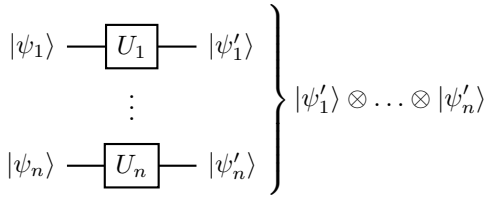
This expression has the same form as Equation (5.5.1) with $|\phi_i\rangle = \alpha_i|b\rangle$, i.e. each of the $|\phi_i\rangle$ vectors in this expansion is a multiple of the same vector $|b\rangle$. Conversely, if $|\phi_i\rangle = \alpha_i|b\rangle$ for all i in Equation (5.5.1), then $|\psi\rangle$ must be a product state. So if we want to identify which joint states are product states and which are not, we simply write the joint state according to Equation (5.5.1) and check if all the vectors $|\phi_i\rangle$ are multiples of a single vector. Needless to say, if we choose the states $|\phi\rangle$ randomly, it is very unlikely that this condition is satisfied, and we almost certainly pick an entangled state.

Even though an entangled state cannot be written as a tensor product, it can always be written as a linear combination of vectors from the tensor product basis. In fact, any state of n qubits $|\psi\rangle$ can be expressed in the standard *product* basis. In general, given n qubits, we need $2(2^n - 1)$ real parameters to describe their state vector, but only $2n$ to describe separable states.

Quantum entanglement is one of the most fascinating aspects of quantum theory. We will now explore some of its implications.

5.6 Controlled-NOT

How do entangled states arise in real physical situations? The short answer is that *entanglement is the result of interactions*. It is easy to see that tensor product operations $U_1 \otimes \ldots U_n$ map product states to product states:



and so any collection of separable qubits remains separable.

As soon as qubits start interacting with one another, however, they become entangled, and things start to get really interesting. We will describe interactions that cannot be written as tensor products of unitary operations on individual qubits. The most popular two-qubit entangling gate is the **controlled-NOT** (or `c-NOT`), also known as the **controlled-X** gate. The gate acts on two qubits: it flips the second qubit (referred to as the **target**) if the first qubit (referred to as the **control**) is $|1\rangle$, and does nothing if the control qubit is $|0\rangle$. In the standard basis $\{|00\rangle, |01\rangle, |10\rangle, |11\rangle\}$, it is represented by the following unitary matrix:

Here the X refers to the Pauli operator $\sigma_x \equiv X$ that implements the bit-flip.

controlled-NOT

$$\left[\begin{array}{cc|cc} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ \hline 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{array}\right]$$

We can also represent the c-NOT gate using the circuit notation, as in Figure 27.

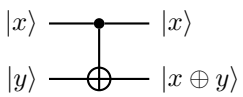


Figure 27. Where $x, y \in \{0, 1\}$, and $\oplus$ denotes XOR, or addition modulo 2.

Note that this gate does not admit any tensor product decomposition, but can be written as a sum of tensor products:

$$\texttt{c-NOT} = |0\rangle\langle 0| \otimes \mathbf{1} + |1\rangle\langle 1| \otimes X$$

where X is the Pauli bit-flip operation.

The c-NOT gate lets us do many interesting things, and can act in a rather deceptive way. Let us now study some of these things.

Make sure that you understand how the Dirac notation is used here. More generally, think why

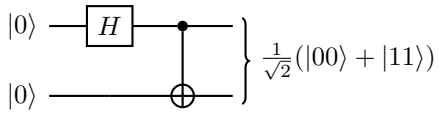
$$|0\rangle\langle 0| \otimes A + |1\rangle\langle 1| \otimes B$$

means "*if the first qubit is in state* $|0\rangle$ *then apply* A *to the second one, and if the first qubit is in state* $|1\rangle$ *then apply* B *to the second one*". What happens if the first qubit is in a superposition of $|0\rangle$ and $|1\rangle$?

5.6.1 The Bell states, and the Bell measurement

We start with the generation of entanglement. Here is a simple circuit that demonstrates the entangling power of c-NOT:

Circuit. (Generating entanglement).



In this circuit, the separable input $|0\rangle|0\rangle$ evolves as

$$\begin{aligned} |0\rangle|0\rangle &\overset{H}{\longmapsto} \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)|0\rangle \\ &= \frac{1}{\sqrt{2}}|0\rangle|0\rangle + \frac{1}{\sqrt{2}}|1\rangle|0\rangle \\ &\overset{\texttt{c-NOT}}{\longmapsto} \frac{1}{\sqrt{2}}|0\rangle|0\rangle + \frac{1}{\sqrt{2}}|1\rangle|1\rangle \end{aligned}$$

resulting in the entangled output $\frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$. In fact, this circuit implements the unitary operation which maps the standard computational basis into the four entangled states, known as the **Bell states**.

John Stewart Bell (1928–1990) was a Northern Irish physicist.

The **Bell states**, $|\psi_{ij}\rangle$, as generated by the above circuit:

$$|00\rangle \mapsto |\psi_{00}\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$$
$$|01\rangle \mapsto |\psi_{01}\rangle = \frac{1}{\sqrt{2}}(|01\rangle + |10\rangle)$$
$$|10\rangle \mapsto |\psi_{10}\rangle = \frac{1}{\sqrt{2}}(|00\rangle - |11\rangle)$$
$$|11\rangle \mapsto |\psi_{11}\rangle = \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle)$$

The more standard notation for these states, however, is the following:

$$\Phi^+ = |\psi_{00}\rangle$$
$$\Psi^+ = |\psi_{01}\rangle$$
$$\Phi^- = |\psi_{10}\rangle$$
$$\Psi^- = |\psi_{11}\rangle$$

(and this is the notation that we will use from now on).

The Bell states form an orthonormal basis in the Hilbert space $\mathcal{H}_1 \otimes \mathcal{H}_2$ of two qubits. We can perform measurements in the Bell basis: the easiest way to do it in practice is to "rotate" the Bell basis to the standard basis, and then perform the measurement in the standard basis. Indeed, if we reverse the circuit, then we get a circuit which maps the Bell state $|\psi_{xy}\rangle$ to the corresponding state $|xy\rangle$ in the standard basis. This unitary mapping allows us to "implement" the projections on Bell states by applying the reversed circuit followed by the usual qubit-by-qubit measurement in the standard basis.

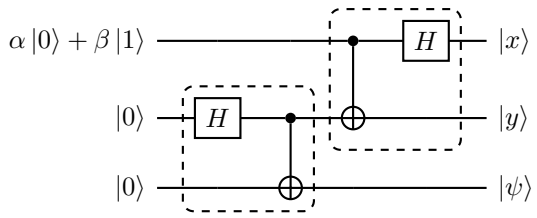
For any state $|\psi\rangle$ of two qubits, the amplitude $\langle\psi_{xy}|\psi\rangle$ can be written as $\langle xy|U^\dagger|\psi\rangle$, where $U^\dagger$ is such that $|\psi_{xy}\rangle = U|xy\rangle$.

5.6.2 Quantum teleportation

A wonderful fact, that sounds more like science fiction than actual science, is the following: *an unknown quantum state can be teleported from one location to another.* Consider the following circuit:

Divide et impera, or "divide and conquer": a good approach to solving problems in mathematics (and in life). Start with the smaller circuits in the dashed boxes.

Circuit. (Quantum teleportation).



The first input qubit (counting from the top) is in some arbitrary state. After the action of the part of the circuit in the first dashed box (counting from the left), the state of the three qubits reads

We neglect to write the normalisation factors.

$$\big(\alpha|0\rangle + \beta|1\rangle\big)\big(|00\rangle + |11\rangle\big).$$

By regrouping the terms, but keeping the qubits in the same order, this state can be written as the sum

$$\begin{aligned}&(|00\rangle + |11\rangle) \otimes (\alpha|0\rangle + \beta|1\rangle)\\ +&(|01\rangle + |10\rangle) \otimes (\alpha|1\rangle + \beta|0\rangle)\\ +&(|00\rangle - |11\rangle) \otimes (\alpha|0\rangle - \beta|1\rangle)\\ +&(|01\rangle - |10\rangle) \otimes (\alpha|1\rangle - \beta|0\rangle).\end{aligned}$$

Then the part of the circuit in the second dashed box maps the four Bell states of the first two qubits to the corresponding states from the computational basis:

$$\begin{aligned}&|00\rangle \otimes (\alpha|0\rangle + \beta|1\rangle)\\ +&|01\rangle \otimes (\alpha|1\rangle + \beta|0\rangle)\\ +&|10\rangle \otimes (\alpha|0\rangle - \beta|1\rangle)\\ +&|11\rangle \otimes (\alpha|1\rangle - \beta|0\rangle).\end{aligned}$$

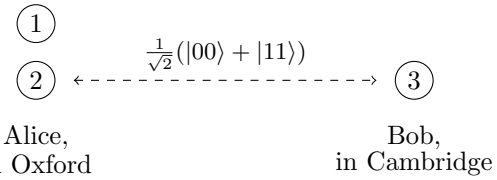
Upon performing the standard measurement and learning the values of x and y, we choose one of the four following transformations depending on these values:

$$\begin{array}{ll} 00 \mapsto \mathbf{1} & 01 \mapsto X \\ 10 \mapsto Z & 11 \mapsto ZX \end{array} \tag{5.6.2}$$

(e.g. if $x = 0$ and $y = 1$, then we choose X). We then apply this transformation to the third qubit, which restores the original state of the first qubit.

If you understand how this circuit works then you are ready for quantum teleportation. Here is a dramatic version.

> *Suppose three qubits, which look very similar, are initially in the possession of an absent-minded Oxford student, Alice. The first qubit is in a precious quantum state and this state is needed urgently for an experiment in Cambridge. The other two qubits are entangled, in the $|\psi_{00}\rangle$ state. Alice's colleague, Bob, pops in to collect the qubit. Once he is gone, Alice realises that, by mistake, she gave him not the first but the third qubit: the one which is entangled with the second qubit.*



> *The situation seems to be hopeless — Alice does not know the quantum state of the first qubit, and Bob is now miles away and her communication with him is limited to few bits. However, Alice and Bob are both very clever and they both diligently attended their "Introduction to Quantum Information Science" classes. Can Alice rectify her mistake and save Cambridge science?*
>
> *Hmmm... (pause for thought)...*
>
> *Of course: Alice can teleport the state of the first qubit! She performs the Bell measurement on the first two qubits, which gives her two binary digits, x and y. She*

then broadcasts x and y to Bob, who chooses the corresponding transformation, as in Equation (5.6.2), performs it, and recovers the original state.

This raises a natural "philosophical" question: what do we really *mean* by teleportation? A key part of this question is understanding what happens to our original qubit when we teleport it. As it turns out, it must necessarily be *destroyed*, as we now explain.

5.6.3 Thou shalt not clone

Let us now look at something that controlled-NOT *seems* to be doing but, in fact, *isn't*. It is easy to see that the c-NOT can copy the bit value of the first qubit:

$$|x\rangle|0\rangle \overset{\texttt{c-NOT}}{\longmapsto} |x\rangle|x\rangle \qquad (\text{for } x = 0, 1)$$

so one might suppose that this gate could also be used to copy superpositions, such as $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$, so that

$$|\psi\rangle|0\rangle \overset{\texttt{c-NOT}}{\longmapsto} |\psi\rangle|\psi\rangle$$

for any $|\psi\rangle$.

This is not so!

The unitarity of the c-NOT means that it turns superpositions in the control qubit into *entanglement* of the control and the target: if the control qubit is in the a superposition state $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$ (with $\alpha, \beta \neq 0$), and the target is in $|0\rangle$, then the c-NOT gate generates the entangled state

$$\big(\alpha|0\rangle + \beta|1\rangle\big)|0\rangle \overset{\texttt{c-NOT}}{\longmapsto} \alpha|00\rangle + \beta|11\rangle.$$

In fact, it is *impossible* to clone an unknown quantum state, and we can prove this!

To prove this via contradiction, let's assume that we *could* build a universal quantum cloner, and then take any two normalised states $|\psi\rangle$ and $|\phi\rangle$ that are *non-identical* (i.e. $|\langle\psi|\phi\rangle| \neq 1$) and *non-orthogonal* (i.e. $\langle\psi|\phi\rangle \neq 0$). If we then run our hypothetical cloning machine we get

$$|\psi\rangle|0\rangle|W\rangle \mapsto |\psi\rangle|\psi\rangle|W'\rangle$$
$$|\phi\rangle|0\rangle|W\rangle \mapsto |\phi\rangle|\phi\rangle|W''\rangle$$

where the third system, initially in state $|W\rangle$, represents everything else (say, the internal state of the cloning machine). For this transformation to be unitary, it must preserve the inner product, and so we require that

$$\langle\psi|\phi\rangle = \langle\psi|\phi\rangle^2\langle W'|W''\rangle$$

which can only be satisfied if $|\langle\psi|\phi\rangle|$ is equal to 1 or 0, but this contradicts our assumptions!

Thus, states of qubits, unlike states of classical bits, cannot be faithfully cloned. Note that, in quantum teleportation, the original state must therefore be *destroyed*, since otherwise we would be producing a clone of an unknown quantum state. The no-cloning property of quantum states leads to interesting applications, of which quantum cryptography is one.

Universal quantum cloners are *impossible*.

5.7 Other controlled gates

5.7.1 Controlled-phase

Needless to say, not everything is about the controlled-NOT gate. Another common two-qubit gate is the **controlled-phase** gate $\texttt{c-}P_\varphi$.

controlled-phase

$$\left[\begin{array}{cc|cc} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ \hline 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & e^{i\varphi} \end{array}\right]$$

We can also represent the $\texttt{c-}P_\varphi$ gate using the circuit notation, as in Figure 28.

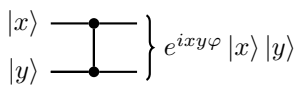
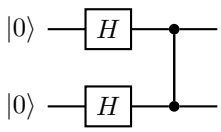


Figure 28. Where $x, y \in \{0, 1\}$.

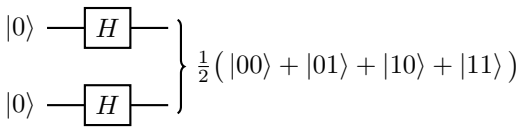
Again, the matrix is written in the computational basis $\{|00\rangle, |01\rangle, |10\rangle, |11\rangle\}$. If we do not specify the phase then we usually assume that $\varphi = \pi$, in which case we call this operation the **controlled-Z gate**, which acts as $|0\rangle\langle 0|0 \otimes \mathbf{1} + |1\rangle\langle 1|1 \otimes Z$. Here Z refers again to the Pauli phase-flip $\sigma_z \equiv Z$ operation.

In order to see the entangling power of the controlled-phase shift gate, consider the following circuit.

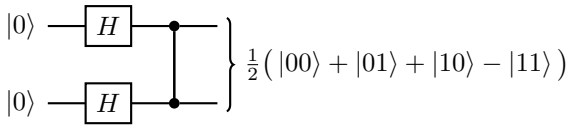
Circuit. (Generating entanglement, again).



In this circuit, first the two Hadamard gates prepare the equally-weighted superposition of all states from the computational basis



and then the controlled-Z operation flips the sign in front of $|11\rangle$



which results in the entangled state.

5.7.2 Controlled-U

Both the above two-qubit controlled gates (i.e. `c-NOT` and `c-`P_φ) are specific examples of the more general construction of a **controlled-U gate**:

$$\texttt{c-}U = |0\rangle\langle 0|0 \otimes \mathbf{1} + |1\rangle\langle 1|1 \otimes U$$

where U is an arbitrary single-qubit unitary transformation U.

controlled-U
$\left[\begin{array}{cc\|cc} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ \hline 0 & 0 & \multicolumn{2}{c}{\multirow{2}{*}{U}} \\ 0 & 0 & & \end{array}\right]$

We can also represent the `c-`U gate using the circuit notation, as in Figure 29.

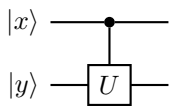


Figure 29. Where $x, y \in \{0, 1\}$.

We can go even further and consider a more general unitary operation: the two-qubit **x-controlled-U gate**:

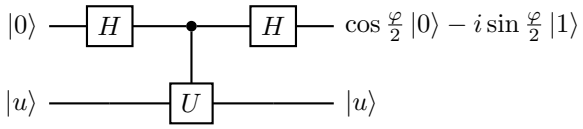
$$\sum_x |x\rangle\langle x| \otimes U_x \equiv |0\rangle\langle 0| \otimes U_0 + |1\rangle\langle 1| \otimes U_1$$

where each U_x is a unitary transformation that is applied to the second qubit only if the first one is in state $|x\rangle$. In general, an x-controlled-U gate can be defined on two registers of arbitrary size n and m, with $x \in \{0, 1\}^n$ and the U_x being $(2^m \times 2^m)$ unitary matrices acting on the second register.

5.7.3 Phase kick-back

Before moving on to the next section, we first describe a simple "trick" — an unusual way of introducing phase shifts that will be essential for our analysis of quantum algorithms. Consider the following circuit.

Circuit. (Controlled-U interference).



where $|u\rangle$ is an *eigenstate* of U (that is, $U|u\rangle = e^{i\varphi}|u\rangle$ for some φ).

This should look familiar: it is the usual interference circuit, but with the phase gate replaced by a controlled-U gate, which will mimic the phase gate, as we shall soon see. Note that the second qubit is prepared in state $|u\rangle$, which is *required to be an eigenstate of* U. The circuit effects the following sequence of transformations (omitting the normalisation factors):

$$\begin{aligned}
|0\rangle|u\rangle &\overset{H}{\longmapsto} (|0\rangle + |1\rangle)|u\rangle \\
&= |0\rangle|u\rangle + |1\rangle|u\rangle \\
&\overset{\text{c-}U}{\longmapsto} |0\rangle|u\rangle + |1\rangle U|u\rangle \\
&= |0\rangle|u\rangle + e^{i\varphi}|1\rangle|u\rangle \\
&= (|0\rangle + e^{i\varphi}|1\rangle)|u\rangle \\
&\overset{H}{\longmapsto} \left(\cos\frac{\varphi}{2}|0\rangle - i\sin\frac{\varphi}{2}|1\rangle\right)|u\rangle.
\end{aligned}$$

Note that the second qubit does *not* get entangled with the first one: it remains in its original state $|u\rangle$. However, the interaction between the two qubits introduces a phase shift on the first qubit. This may look like an unnecessarily complicated way of introducing phase shifts, but, as we shall soon see, this is how quantum computers do it. Let me give you a preview of things to come.

Consider the following x-controlled-U operation:

$$\begin{bmatrix}
\mathbf{1} & 0 & 0 & 0 \\
0 & \mathbf{1} & 0 & 0 \\
0 & 0 & \mathbf{1} & 0 \\
0 & 0 & 0 & X
\end{bmatrix}$$

$$\begin{aligned}
&= |00\rangle\langle 00| \otimes \mathbf{1} \\
&+ |01\rangle\langle 01| \otimes \mathbf{1} \\
&+ |10\rangle\langle 10| \otimes \mathbf{1} \\
&+ |11\rangle\langle 11| \otimes X.
\end{aligned}$$

The first register is of size 2, and the second register is of size 1. If the first register is prepared in state $|11\rangle$, then the qubit in the second register is flipped (by the Pauli bit-flip X); otherwise, nothing happens. This unitary operation is a quantum version of the

Boolean function evaluation: it corresponds to the Boolean function

$$\begin{aligned} f\colon \{0,1\}^2 &\to \{0,1\} \\ 00 &\mapsto 0 \\ 01 &\mapsto 0 \\ 10 &\mapsto 0 \\ 11 &\mapsto 1. \end{aligned}$$

If $f(x) = 1$, then we flip the bit value in the second register (with operation X); if $f(x) = 0$, then we do nothing.

Now, prepare the qubit in the second register in state $|0\rangle - |1\rangle$, which is an eigenstate of X with eigenvalue $e^{\pi i} = -1$. So whenever X is applied to the second register, the phase factor -1 appears in front of the corresponding term in the first register. If we prepare the first register in the superposition $|00\rangle + |01\rangle + |10\rangle + |11\rangle$ then the result of applying the above x-controlled-U operation is the entangled state $|00\rangle + |01\rangle + |10\rangle - |11\rangle$. That is, *the phase kick-back mechanism introduced a relative phase in the equally-weighted superposition of all binary strings of length two*.

> Phase kick-back is how we control quantum interference in quantum computation.

We will return to this topic later on, when we discuss quantum evaluation of Boolean functions and quantum algorithms.

5.7.4 Universality revisited

We will come across few more gates in this course, but at this stage you already know all the elementary unitary operations that are needed to construct any unitary operation on any number of qubits:

- the Hadamard gate,
- all phase gates, and
- the `c-NOT`

These gates form a **universal set of gates**: with $O(4^n n)$ of these gates, we can construct any n-qubit unitary operation. We should mention that there are many universal sets of gates. In fact, almost any gate that can entangle two qubits can be used as a universal gate.

Recall the big-O asymptotic notation: given a *positive* function $f(n)$, we write $O(f(n))$ to mean "bounded above by $c\,f(n)$ for some constant $c > 0$ (for sufficiently large n)". For example, $15n^2 + 4n + 7$ is $O(n^2)$.

We are particularly interested in any *finite* universal set of gates (such as the one containing the Hadamard, $P_{\frac{\pi}{4}}$ (the T gate), and the `c-NOT`), which can approximate any unitary operation on n qubits with arbitrary precision. The price to pay is the number of gates — better precision requires more gates. We shall elaborate on this later.

5.7.5 Density operators and the like

The existence of entangled states leads to an obvious question: if we cannot attribute a state vector to an individual qubit, then how can we describe its quantum state? In the next few chapters we will see that, when we limit our attentions to a part of a larger system, states are not represented by vectors, measurements are not described by orthogonal projections, and evolution is not unitary. As a spoiler, here is a dictionary of some of the new concepts that will soon be introduced:

state vectors	$\longmapsto$	density operators
unitary evolutions	$\longmapsto$	completely-positive trace-preserving maps
orthogonal projectors	$\longmapsto$	positive operator-valued measures

5.8 Why qubits, subsystems, and entanglement?

One question that is rather natural to ask at this point is the following:

> *If entanglement is so fragile and difficult to control, then why bother? Why not perform the whole computation in one physical system that has as many quantum states as we normally have labels for the states of qubits, so that we can label these quantum states in the same way as we normally label the qubits, and give* them *computational meaning?*

This suggestion, although possible, gives a very inefficient way of representing data (it is describing the **unary encoding**). For serious computations *we need subsystems*. Here is why.

Suppose you have n physical objects, and each object has k distinguishable states. If you can access each object *separately* and put it into any of the k states, then, with only n operations, you can prepare any of the k^n different configurations of the combined systems. Without any loss of generality, let us take $k = 2$ and refer to each object of this type as a **physical bit**. We label the two states of a physical bit as 0 and 1. Any collection of n physical bits can be prepared in 2^n different configurations, which can be used to store up to 2^n messages/binary strings/different numbers. In order to represent numbers from 0 to $N-1$ we just have to choose n such that $N \leqslant 2^n$.

Suppose the two states in the physical bit are separated by the energy difference ΔE. Then a preparation of any particular configuration will cost no more than $E = n\Delta E = (\log_2 N)\Delta E$ units of energy.

In contrast, if we choose to encode N configurations into one chunk of matter, say, into the first N energy states of a single harmonic oscillator with the interstate energy separation ΔE then, in the worst case, one has to use $E = N\Delta E$ units of energy, e.g. to go from the ground state (labelled as 0) to the most excited state (labelled as N). For large N this gives an exponential gap in the energy expenditure between the binary encoding using physical bits and unary encoding, using energy levels of harmonic oscillators.

One can, of course, try to switch from harmonic oscillators to quantum systems which have a finite spread in the energy spectrum. For example, by operating on the energy states of the hydrogen atom one can encode any number from 0 to $N-1$, and one is guaranteed not to spend more than $E_{\max} = 13.6\,\mathrm{eV}$ (otherwise the atom is ionised). The snag is that, in this case, some of the electronic states will be separated by the energy difference to the order of $E_{\max}/N$, and to drive the system selectively from one state to another one has to tune into the frequency $E_{\max}/\hbar N$, which requires a sufficiently long wave-packet (so that the frequency is well defined), and consequently the interaction time is of order $N(\hbar/E_{\max})$.

That is, we have to trade energy for time.

It turns out that whichever way we try to represent the number N using the unary encoding (i.e. using N different states of a single chunk of matter), we end up depleting our physical resources (such as energy, time, space) at a much greater rate than in the case when we use subsystems. This plausibility argument indicates that, for efficient processing of information, the system must be divided into subsystems — for example, into physical bits.

5.9 Remarks and exercises

5.9.1 Entangled or not?

Let a joint state of $\mathcal{A}$ and $\mathcal{B}$ be written in a product basis as

$$|\psi\rangle = \sum_{ij} c_{ij}|a_i\rangle \otimes |b_j\rangle.$$

Assume that $\mathcal{H}_a$ and $\mathcal{H}_b$ are of the same dimension.

- Show that, if $|\psi\rangle$ is a product state, then $\det(c_{ij}) = 0$.
- Show that the converse ($\det(c_{ij}) = 0$ implies that $|\psi\rangle$ is a product state) holds only for qubits. Explain why.
- Deduce that the state

  $$\frac{1}{2}\big(|00\rangle + |01\rangle + |10\rangle + (-1)^k|11\rangle\big)$$

  is entangled for $k = 1$ and unentangled for $k = 0$. Express the latter case explicitly as a product state.

There is a lot of interesting physics behind the previous innocuous-looking mathematical statement. For example, think again about the state $(|00\rangle + |11\rangle)/\sqrt{2}$. What happens if you measure just the first qubit? It is equally likely that you get $|0\rangle$ or $|1\rangle$, right? But after your measurement the two qubits are either in state $|00\rangle$ or in $|11\rangle$, i.e. they show the same bit value. Now, why might that be disturbing? Well, imagine the second qubit to be light-years away from the first one. It seems that the measurement of the first qubit affects the second qubit right away, which seems to imply faster-than-light communication! This is what Einstein called "spooky action as a distance". But can you actually use this effect to send a message faster than light? What would happen if you tried?

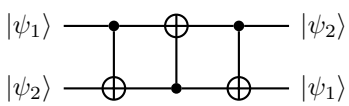
"Spooky action at a distance" is a loose translation of the German "spukhafte Fernwirkung", which is the phrase that Albert Einstein used in his 1947 letter to Max Born.

Hopefully you can see that it would not work, since the result of the measurement is random — you cannot choose the bit value you want to send. We shall return to this, and other related phenomena, later on.

5.9.2

Show that, for any states $|\psi_1\rangle$ and $|\psi_2\rangle$, the circuit below implements the swap operation $|\psi_1\rangle|\psi_2\rangle \mapsto |\psi_2\rangle|\psi_1\rangle$.

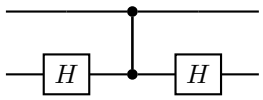
Circuit. (Swapping).



5.9.3

Show that the circuit below implements the controlled-`NOT` gate.

Circuit. (Controlled-NOT, again).

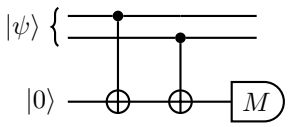


5.9.4

The controlled-NOT gate can act as a measurement gate: if you prepare the target in state $|0\rangle$ then the gate acts as $|x\rangle|0\rangle \mapsto |x\rangle|x\rangle$, and so the target learns the bit value of the control qubit. If you wish, you can think about a subsequent measurement of the target qubit in the computational basis as an observer learning about the bit value of the control qubit.

Take a look at the circuit below, where M stands for measurement in the standard basis.

Circuit. (?).



Now assume that the top two qubits are in the state

$$|\psi\rangle = \frac{1}{\sqrt{3}}\big(|01\rangle - |10\rangle + i|11\rangle\big).$$

The measurement M gives two outcomes: 0 and 1. What are the probabilities of each outcome, and what is the post-measurement state in each case? Further, what is actually being measured here?

5.9.5 Arbitrary controlled-U on two qubits

Any unitary operation U on a single qubit can be expressed as

$$U = B^\dagger X B A^\dagger X A$$

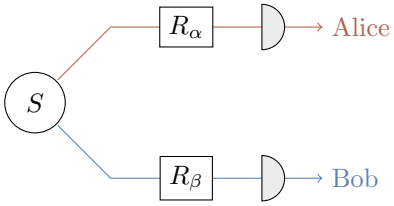
where $X \equiv \sigma_x$ is the Pauli bit-flip operator, and A and B are unitaries. Suppose that you can implement any single qubit gate, and that you have a bunch of controlled-NOT gates at your disposal. How would you implement any controlled-U operation on two qubits?

5.9.6 Entangled qubits

Two entangled qubits in the state $\frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$ are generated by some source S. One qubit is sent to Alice, and one to Bob, who then both perform measurements in the computational basis.

1. What is the probability that Alice and Bob will register identical results? Can any correlations they observe be used for instantaneous communication?

2. Prior to the measurements in the computational basis, Alice and Bob apply unitary operations R_α and R_β (respectively) to their respective qubits:



The gate R_θ is defined by its action on the basis states:

$$|0\rangle \longmapsto \cos\theta|0\rangle + \sin\theta|1\rangle$$
$$|1\rangle \longmapsto -\sin\theta|0\rangle + \cos\theta|1\rangle.$$

Show that the state of the two qubits prior to the measurements is

$$\frac{1}{\sqrt{2}}\cos(\alpha-\beta)\big(|00\rangle + |11\rangle\big)$$
$$-\frac{1}{\sqrt{2}}\sin(\alpha-\beta)\big(|01\rangle - |10\rangle\big).$$

3. What is the probability that Alice and Bob's outcomes are identical?

5.9.7 Quantum dense coding

!!!TODO!!!

5.9.8 Playing with conditional unitaries

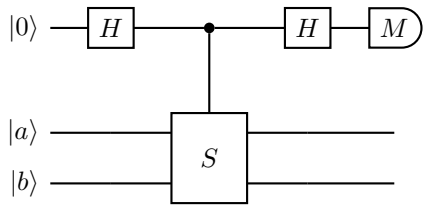
The swap gate S on two qubits is defined first on product vectors by $S\colon |a\rangle|b\rangle \mapsto |b\rangle|a\rangle$ and then extended to sums of product vectors by linearity.

1. Show that the four Bell states $\frac{1}{\sqrt{2}}(|00\rangle \pm |11\rangle)$ and $\frac{1}{\sqrt{2}}(|01\rangle \pm |10\rangle)$ are eigenvectors of S that form an orthonormal basis in the Hilbert space associated to two qubits. Which Bell states span the *symmetric subspace* (i.e. the space spanned by all eigenvectors with eigenvalue 1), and which the *antisymmetric* one (i.e. that spanned by eigenvectors with eigenvalue -1)? Can S have any eigenvalues apart from ± 1?

2. Show that $P_\pm = \frac{1}{2}(\mathbf{1} \pm S)$ are two orthogonal projectors which form the decomposition of the identity and project on the symmetric and antisymmetric subspaces. Decompose the state vector $|a\rangle|b\rangle$ of two qubits into symmetric and antisymmetric components.

3. Consider the quantum circuit below, composed of two Hadamard gates, one controlled-S operation (also known as the **controlled-swap**, or **Fredkin** gate), and the measurement M in the computational basis. Suppose that the state vectors $|a\rangle$ and $|b\rangle$ are normalised but *not* orthogonal to one another. Step through the execution of this network, writing down the quantum states of the three qubits after each computational step. What are the probabilities of observing 0 or 1 when the measurement M is performed?

Circuit. (Symmetric and antisymmetric projection).



4. Explain why this quantum network implements projections on the symmetric and antisymmetric subspaces of the two qubits.

5. Two qubits are transmitted through a quantum channel which applies the same, randomly chosen, unitary operation U to each of them. Show that the symmetric and antisymmetric subspaces are invariant under $U \otimes U$.

6. Polarised photons are transmitted through an optical fibre. Due to the variation of the refractive index along the fibre, the polarisation of each photon is rotated by the same unknown angle. This makes communication based on polarisation encoding unreliable. However, if you can prepare any polarisation state of the two photons then you can still use the channel to communicate without any errors. How can this be achieved?

5.10 Appendix: Tensor products in components

In our discussion of tensor products we have so far taken a rather abstract approach. There are, however, situations in which we have to put numbers in, and write tensor products of vectors and matrices explicitly. For example, here is the standard basis of two qubits written explicitly as column vectors:

$$|00\rangle \equiv |0\rangle \otimes |0\rangle = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \otimes \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

$$|01\rangle \equiv |0\rangle \otimes |1\rangle = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \otimes \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}$$

$$|10\rangle \equiv |1\rangle \otimes |0\rangle = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \otimes \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}$$

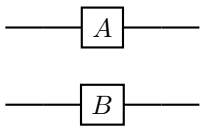
$$|11\rangle \equiv |1\rangle \otimes |1\rangle = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \otimes \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

Given $|a\rangle = \alpha_0|0\rangle + \alpha_1|1\rangle$ and $|b\rangle = \beta_0|0\rangle + \beta_1|1\rangle$, we write $|a\rangle \otimes |b\rangle$ as

$$\begin{aligned}|a\rangle \otimes |b\rangle &= \begin{bmatrix}\alpha_0\\ \alpha_1\end{bmatrix} \otimes \begin{bmatrix}\beta_0\\ \beta_1\end{bmatrix} \\ &= \begin{bmatrix}\alpha_0 \begin{bmatrix}\beta_0\\ \beta_1\end{bmatrix} \\ \alpha_1 \begin{bmatrix}\beta_0\\ \beta_1\end{bmatrix}\end{bmatrix} \\ &= \begin{bmatrix}\alpha_0\beta_0\\ \alpha_0\beta_1\\ \alpha_1\beta_0\\ \alpha_1\beta_1\end{bmatrix}.\end{aligned}$$

Note that each element of the first vector multiplies the entire second vector. This is often the easiest way to get the tensor products in practice.

The matrix elements of the tensor product operation $A \otimes B$



are given by

$$(A \otimes B)_{ik,jl} = A_{ij}B_{kl}$$

where $ik \in \{00, 01, 10, 11\}$ labels the rows, and $kl \in \{00, 01, 10, 11\}$ labels columns, when forming the block matrix:

We always use the lexicographic order $00 < 01 < 10 < 11$.

$$\begin{aligned}A \otimes B &= \begin{bmatrix}A_{00} & A_{01}\\ A_{10} & A_{11}\end{bmatrix} \otimes \begin{bmatrix}B_{00} & B_{01}\\ B_{10} & B_{11}\end{bmatrix} \\ &= \begin{bmatrix}A_{00}B & A_{01}B\\ A_{10}B & A_{11}B\end{bmatrix} \\ &= \left[\begin{array}{cc|cc} A_{00}B_{00} & A_{00}B_{01} & A_{01}B_{00} & A_{01}B_{01}\\ A_{00}B_{10} & A_{00}B_{11} & A_{01}B_{10} & A_{01}B_{11}\\ \hline A_{10}B_{00} & A_{10}B_{01} & A_{11}B_{00} & A_{11}B_{01}\\ A_{10}B_{10} & A_{10}B_{11} & A_{11}B_{10} & A_{11}B_{11}\end{array}\right]\end{aligned}$$

The tensor product induces a natural partition of matrices into blocks. Multiplication of block matrices works pretty much the same as regular matrix multiplication (assuming the dimensions of the sub-matrices are appropriate), except that the entries are now matrices rather than numbers, and so may not commute.

1. Evaluate the following matrix product of (4×4) block matrices:

$$\left[\begin{array}{c|c} \mathbf{1} & X\\ \hline Y & Z\end{array}\right] \left[\begin{array}{c|c} \mathbf{1} & Y\\ \hline X & Z\end{array}\right]$$

(where X, Y, and Z are the Pauli matrices).

2. Using the block matrix form of $A \otimes B$ expressed in terms of A_{ij} and B_{ij} (as described above), explain how the following operations are performed on the block matrix:

   - transposition $(A \otimes B)^T$; partial transpositions $A^T \otimes B$ and $A \otimes B^T$;
   - trace $\mathrm{tr}(A \otimes B)$; partial traces $(\mathrm{tr}\, A) \otimes B$ and $A \otimes (\mathrm{tr}\, B)$.

   Consider the Hadamard transform $H \otimes H \otimes H$ on three qubits, which is described by a $(2^3 \times 2^3)$ matrix. We know that

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

   and so we can calculate that

   $$H \otimes H = \frac{1}{2} \left[\begin{array}{cc|cc} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ \hline 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{array}\right]$$

   and thus that

   $$H \otimes H \otimes H = \frac{1}{2} \left[\begin{array}{cc|cc|cc|cc} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \\ \hline 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 \\ \hline 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 \\ \hline 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 \\ 1 & -1 & -1 & 1 & -1 & 1 & 1 & -1 \end{array}\right].$$

   The rows and columns of $H \otimes H \otimes H$ are labelled by the triples $000, 001, \ldots, 111$. Now, suppose we apply $H \otimes H \otimes H$ to the state $|110\rangle$:

   $$\left.\begin{array}{l} |1\rangle \text{—}\boxed{H}\text{—} \frac{|0\rangle - |1\rangle}{\sqrt{2}} \\ |1\rangle \text{—}\boxed{H}\text{—} \frac{|0\rangle - |1\rangle}{\sqrt{2}} \\ |0\rangle \text{—}\boxed{H}\text{—} \frac{|0\rangle + |1\rangle}{\sqrt{2}} \end{array}\right\} = \frac{1}{2^{3/2}} \begin{pmatrix} |000\rangle + |001\rangle - |010\rangle - |011\rangle \\ -|100\rangle - |101\rangle + |110\rangle + |111\rangle \end{pmatrix}$$

3. The output state is a superposition of all binary strings: $\sum_x c_x |x\rangle$, with $x \in \{0,1\}^3$. Where in the $H^{\otimes 3}$ matrix will you find the coefficients c_x?

   Now, do you want to write down $H \otimes H \otimes H \otimes H$? I don't think so. This is an exponentially growing monster and you may soon run out of space if you actually do try to write it down. Instead, let's spot the pattern of the entries ± 1 in these matrices.

4. Consider the Hadamard gate matrix H_{ab}, where $a, b = 0, 1$ are the labels for the rows and the columns. Observe that $H_{ab} = (-1)^{ab}/\sqrt{2}$. (This may look like a fancy way of writing the entries of the Hadamard matrix but it will pay off in a moment). Using the fact that $(A \otimes B)_{ik,jl} = A_{ij} B_{kl}$, or any other method, analyse the pattern of the ± 1 in the tensor product of Hadamard matrices. What is the entry $H^{\otimes 4}_{0101,1110}$?

5. For any two binary strings $a = (a_1, \ldots, a_n)$ and $b = (b_1, \ldots, b_n)$ of the same length we can define their "scalar" product as $a \cdot b = (a_1 b_1 \oplus \ldots \oplus a_n b_n)$. Show that, up to the constant $(1/\sqrt{2})^n$, the entry $H^{\otimes n}_{a,b}$, for any n and for any binary strings a and b of length n, is $(-1)^{a \cdot b}$.

6. Show that $H^{\otimes n}$ acts as

$$|a\rangle \longmapsto \left(\frac{1}{\sqrt{2}}\right)^n \sum_{b \in \{0,1\}^n} (-1)^{a \cdot b} |b\rangle$$

7. A quantum register of 10 qubits holds the binary string 0110101001. The Hadamard Transform is then applied to this register yielding a superposition of all binary strings of length 10. What is the sign in front of the $|0101010101\rangle$ term?

5.11 Appendix: The Schmidt decomposition

An arbitrary vector in the Hilbert space $\mathcal{H}_\mathcal{A} \otimes \mathcal{H}_\mathcal{B}$ can be expanded in a product basis as

$$|\psi\rangle = \sum_{ij} c_{ij} |a_i\rangle |b_j\rangle.$$

Moreover, for each particular joint state $|\psi\rangle$, we can find orthonormal bases, $\{|\widetilde{a_i}\rangle\}$ in $\mathcal{H}_\mathcal{A}$ and $\{|\widetilde{b_j}\rangle\}$ in $\mathcal{H}_\mathcal{B}$, such that $|\psi\rangle$ can be expressed as

$$|\psi\rangle = \sum_i d_i |\tilde{a}_i\rangle |\tilde{b}_i\rangle,$$

where the coefficients d_i are non-negative numbers. This is known as the **Schmidt decomposition** of $|\psi\rangle$. Any bipartite state can be expressed in this form, but remember that *the bases used depend on the state being expanded*. Indeed, given two bipartite states $|\psi\rangle$ and $|\phi\rangle$, we usually *cannot* perform the Schmidt decomposition using the *same* orthonormal bases in $\mathcal{H}_\mathcal{A}$ and $\mathcal{H}_\mathcal{B}$. The number of terms in the Schmidt decomposition is, at most, the minimum of $\dim \mathcal{H}_\mathcal{A}$ and $\dim \mathcal{H}_\mathcal{B}$.

The Schmidt decomposition follows from the **singular value decomposition** or **SVD**): *any* $(n \times m)$ matrix C can be written as

$$C = UDV$$

where U and V are (respectively) $(n \times n)$ and $(m \times m)$ unitary matrices and D is an $(n \times m)$ diagonal matrix with real, non-negative elements in descending order $d_1 \geqslant d_2 \geqslant \ldots \geqslant d_{\min(n,m)}$ (and with the rest of the matrix is filled with zeros). The elements d_k are called the **singular values** of C.

You can visualize the SVD by thinking of C as representing a linear transformation from m-dimensional to n-dimensional Euclidean space: it maps the unit ball in the m-dimensional space to an ellipsoid in the n-dimensional space; the singular values are the lengths of the semi-axes of that ellipsoid; the matrices U and V carry information about the locations of those axes and the vectors in the first space which map into them. Thus SVD tells us that the transformation C is composed of rotating the unit ball (transformation V), stretching the axes by factors d_k, and then rotating the resulting ellipsoid (transformation U).

Using the index notation $C_{ij} = \sum_k U_{ik} d_k V_{kj}$, we can thus apply SVD to c_{ij},

$$\begin{aligned}|\psi\rangle &= \sum_{i,j} c_{ij} |a_i b_j\rangle \\ &= \sum_{i,j} \sum_k U_{ik} d_k V_{kj} |a_i b_j\rangle \\ &= \sum_k d_k \left(\sum_i U_{ik} |a_i\rangle \right) \otimes \left(\sum_j V_{kj} |b_j\rangle \right).\end{aligned}$$

The Schmidt decomposition of a separable state of the form $|a\rangle \otimes |b\rangle$ is trivially just this state. The Bell states Ψ^+ and Φ^+ are already written in their Schmidt form, whereas Ψ^- and Φ^- can be easily expressed in the Schmidt form. For example, for $|\Psi^-\rangle$ we have $d_1 = d_2 = \frac{1}{\sqrt{2}}$, and the Schmidt basis is $|\bar{a}_1\rangle = |0\rangle, |\bar{a}_2\rangle = |1\rangle$, $|\bar{b}_1\rangle = |1\rangle, |\bar{b}_2\rangle = -|0\rangle$. The number of non-zero singular values of c_{ij} is called the **rank** of c_{ij}, or the rank of the corresponding quantum state, or sometimes, the **Schmidt number**. Clearly, all bipartite states of rank one are separable.

The Schmidt decomposition is *almost* unique. The ambiguity arises when we have two or more identical singular values, as, for example, in the case of the Bell states. Then any unitary transformation of the basis vectors corresponding to a degenerate singular value, both in $\mathcal{H}_a$ and in $\mathcal{H}_b$, generates another set of basis vectors.

6 Density matrices

*About **density matrices**, and how they help to solve the problem introduced by entangled states, as well as how they let us talk about mixtures and subsystems. Also a first look at the **partial trace**.*

We cannot always assign a definite state vector to a quantum system. It may be that the system is part of a composite system that is in an entangled state, or it may be that our knowledge of the preparation of a particular system is insufficient to determine its state (for example, someone may prepare a particle in one of the states $|\psi_1\rangle, |\psi_2\rangle, \ldots, |\psi_n\rangle$, with (respective) probabilities $p_1, p_2, \ldots, p_n$). Nevertheless, in either case we are able to make statistical predictions about the outcomes of measurements performed on the system using a more general description of quantum states.

We have already mentioned that the existence of entangled states begs an obvious question: if we cannot attribute a state vectors to an individual quantum system then how shall we describe its quantum state? In this chapter we will introduce an alternate description of quantum states that can be applied both to a composite system and to any of its subsystems. Our new mathematical tool is called a **density operator**. We will start with the density operator as a description of the mixture of quantum states, and will then discuss the partial trace, which is a unique operation that takes care of the reduction of a density operator of a composite system to density operators of its components.

If we choose a particular basis, operators become matrices. Here I will use both terms (density *operators* and density *matrices*) interchangeably.

6.1 Definitions

If you are an impatient mathematically minded person, who feels more comfortable when things are properly defined right from the beginning, here is your definition:

> A **density operator** ρ on a finite dimensional Hilbert space $\mathcal{H}$ is any non-negative self-adjoint operator with trace equal to one.

A self-adjoint matrix M is said to be **non-negative**, or **positive semi-definite**, if $\langle v|M|v\rangle \geqslant 0$ for any vector $|v\rangle$, or if all of its eigenvalues are non-negative, or if here exists a matrix A such that $M = A^\dagger A$. (This is called a **Cholesky factorization**.)

It follows that any such ρ can always be diagonalised, that the eigenvalues are all real and non-negative, and that the eigenvalues sum to one. Moreover, given two density operators ρ_1 and ρ_2, we can always construct another density operator as a convex sum of the two:

$$\rho = p_1\rho_1 + p_2\rho_2 \qquad \text{where} \quad p_1, p_2 \geqslant 0 \text{ and } p_1 + p_2 = 1.$$

You should check that ρ has all the defining properties of a density matrix, i.e. that it is self-adjoint, non-negative, and that its trace is one. This means that density operators form a convex set.

A subset of a vector space is said to be **convex** if, for any two points in the subset, the straight line segment joining them is also entirely contained inside the subset.

The **rank** of a matrix is the number of its non-zero eigenvalues.

An important example of a density operator is a rank one projector. Any quantum state that can be described by the state vector $|\psi\rangle$, called a **pure state**, can be also described by the density operator $\rho = |\psi\rangle\langle\psi|$. Pure states are the extremal points in the convex set of density operators: they cannot be expressed as a convex sum of other elements in the set. In contrast, all other states, called **mixed states**, can be always written as the convex sum of pure states: $\sum_i p_i|\psi_i\rangle\langle\psi_i|$ ($p_i \geqslant 0$ and $\sum_i p_i = 1$). Now that we have cleared the mathematical essentials, we will turn to physical applications.

6.2 Mixtures

Let us start with probability distributions over state vectors. Suppose Alice prepares a quantum system and hands it over to Bob who subsequently measures observable M. If Alice's preparation is described by a state vector $|\psi\rangle$, then, quantum theory declares, the average value of any observable M is given by $\langle\psi|M|\psi\rangle$, which can be also written as

$$\langle M\rangle = \operatorname{tr} M|\psi\rangle\langle\psi|.$$

If M is one of the orthogonal projectors P_k describing the measurement, then the average $\langle P_k\rangle$ is the probability of the outcome k associated with this projector.

This way of expressing the average value makes a clear separation between the contributions from the state preparation and from the choice of the measurement. We have two operators under the trace: one of them, $|\psi\rangle\langle\psi|$, describes the state preparation, and the other one, M, the measurement. Now, suppose Alice prepares the quantum system in one of the states $|\psi_1\rangle, \ldots, |\psi_m\rangle$, choosing state $|\psi_i\rangle$ with probability p_i, and hands the system to Bob without telling him which state was chosen. The possible states $|\psi_i\rangle$ are normalised but need not be orthogonal. We call this situation a **mixture of the states** $|\psi_i\rangle$, or a **mixed state** for short.

> Remember, a mixture of states is very different from a superposition of states: a superposition *always* yields a definite state vector, whereas a mixture does *not*, and so must be described by a density operator.

Bob knows the ensemble of states $|\psi_1\rangle, \ldots, |\psi_m\rangle$ and the corresponding probability distribution $p_1, \ldots, p_m$, and can hence calculate $\langle M\rangle$ as

A pure state can be seen as a special case of a mixed state, where all but one the probabilities p_i equal zero.

$$\begin{aligned}\langle M\rangle &= \sum_i p_i \left(\operatorname{tr} M|\psi_i\rangle\langle\psi_i|\right)\\ &= \operatorname{tr} M \underbrace{\left(\sum_i p_i|\psi_i\rangle\langle\psi_i|\right)}_{\rho}\\ &= \operatorname{tr} M\rho.\end{aligned}$$

Again, we have two operators under the trace: $\rho = \sum_i p_i|\psi_i\rangle\langle\psi_i|$, which pertains to the state preparation, and M, which describes the measurement. We shall call the operator

$$\rho = \sum_i p_i|\psi_i\rangle\langle\psi_i|$$

the **density operator**, since it has all the defining properties of the density operator (the convex sum of rank one projectors). It depends on the constituent states $|\psi_i\rangle$ and their probabilities, and it describes our ignorance about the state preparation.

Once we have ρ we can make statistical predictions: for any observable M we have

$$\langle M\rangle = \operatorname{tr} M\rho.$$

We see that the exact composition of the mixture does not enter this formula: for computing the statistics associated with any observable property of a system, all that matters is the density operator itself, and not its decomposition into the mixture of states. This is important because any given density operator, with the remarkable exception of a pure state, can arise from many different mixtures of pure states. Consider, for example, the

following three scenarios:

1. Alice flips a fair coin. If the result is `Heads` then she prepares the qubit in the state $|0\rangle$, and if the result is `Tails` then she prepares the qubit in the state $|1\rangle$. She gives Bob the qubit without revealing the result of the coin-flip. Bob's knowledge of the qubit is described by the density matrix

$$\frac{1}{2}|0\rangle\langle 0| + \frac{1}{2}|1\rangle\langle 1| = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{bmatrix}.$$

2. Suppose Alice flips a fair coin, as before, but now if the result is `Heads` then she prepares the qubit in the state $|\bar{0}\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$, and if the result is `Tails` then she prepares the qubit in the state $|\bar{1}\rangle = \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle)$. Bob's knowledge of the qubit is now described by the density matrix

$$\begin{aligned} \frac{1}{2}|\bar{0}\rangle\langle\bar{0}| + \frac{1}{2}|\bar{1}\rangle\langle\bar{1}| &= \frac{1}{2}\begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix} + \frac{1}{2}\begin{bmatrix} \frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \end{bmatrix} \\ &= \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{bmatrix}. \end{aligned}$$

3. Suppose Alice picks up any pair of orthogonal states of a qubit and then flips the coin to chose one of them. Any two orthonormal states of a qubit, $|u_1\rangle$, $|u_2\rangle$, form a complete basis, so the mixture $\frac{1}{2}|u_1\rangle\langle u_1| + \frac{1}{2}|u_2\rangle\langle u_2|$ gives $\frac{1}{2}\mathbf{1}$.

As you can see, these three different preparations yield precisely the same density matrix and are hence statistically indistinguishable. In general, two different mixtures can be distinguished (in a statistical sense) if and only if they yield different density matrices. In fact, the optimal way of distinguishing quantum states with different density operators is still an active area of research.

6.3 A few instructive examples, and some less instructive remarks

1. The density matrix corresponding to the state vector $|\psi\rangle$ is the rank one projector $|\psi\rangle\langle\psi|$. Observe that there is no phase ambiguity, since $|\psi\rangle \mapsto e^{i\phi}|\psi\rangle$ leaves the density matrix unchanged, and each $|\psi\rangle$ gives rise to a distinct density matrix.

2. If Alice prepares a qubit in the state $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$ then the corresponding density matrix is the projector

$$|\psi\rangle\langle\psi| = \begin{bmatrix} |\alpha|^2 & \alpha\beta^\star \\ \alpha^\star\beta & |\beta|^2 \end{bmatrix}.$$

3. You are given a qubit and you are told that it was prepared either in state $|0\rangle$ with probability $|\alpha|^2$ or in state $|1\rangle$ with probability $|\beta|^2$. In this case all you can say is that your qubit is in a mixed state described by the density matrix

$$|\alpha|^2|0\rangle\langle 0| + |\beta|^2|1\rangle\langle 1| = \begin{bmatrix} |\alpha|^2 & 0 \\ 0 & |\beta|^2 \end{bmatrix}.$$

   Diagonal density matrices correspond to classical probability distributions on the set of basis vectors.

4. Suppose you want to distinguish between preparations described by the density matrices in examples 2 and 3. Assume that you are given sufficiently many identically prepared qubits described either by the density matrix in example 2 or by the density matrix in example 3. Which of the two measurements would you choose: the measurement in the standard basis $\{|0\rangle, |1\rangle\}$, or the measurement in the basis $\{|\psi\rangle, |\psi_\perp\rangle\}$? One of the two measurements is completely useless. Which one, and why?

5. In general, the diagonal entries of a density matrix describe the probability distributions on the set of basis vectors. They must add up to one, which is why the trace of any density matrix is one. The off-diagonal elements, often called **coherences**, signal departure from the classical probability distribution and quantify the degree to which a quantum system can interfere (we will discuss this in detail later on). The process in which off-diagonal entries (the parameter ϵ in the matrices below) go to zero is called **decoherence**.

$$\begin{bmatrix} |\alpha|^2 & \alpha\beta^\star \\ \alpha^\star\beta & |\beta|^2 \end{bmatrix} \mapsto \begin{bmatrix} |\alpha|^2 & \epsilon \\ \epsilon^\star & |\beta|^2 \end{bmatrix} \mapsto \begin{bmatrix} |\alpha|^2 & 0 \\ 0 & |\beta|^2 \end{bmatrix}$$

   For $\epsilon = \alpha\beta^\star$ we have a pure quantum state ("full interference capability") and for $\epsilon = 0$ we have a classical probability distribution over the standard basis ("no interference capability").

6. Suppose it is equally likely that your qubit was prepared either in state $\alpha|0\rangle + \beta|1\rangle$ or in state $\alpha|0\rangle - \beta|1\rangle$. This means that your qubit is in a mixed state described by the density matrix

$$\frac{1}{2}\begin{bmatrix} |\alpha|^2 & \alpha\beta^\star \\ \alpha^\star\beta & |\beta|^2 \end{bmatrix} + \frac{1}{2}\begin{bmatrix} |\alpha|^2 & -\alpha\beta^\star \\ -\alpha^\star\beta & |\beta|^2 \end{bmatrix} = \begin{bmatrix} |\alpha|^2 & 0 \\ 0 & |\beta|^2 \end{bmatrix}.$$

   You cannot tell the difference between the equally weighted mixture of $\alpha|0\rangle \pm \beta|1\rangle$ and a mixture of $|0\rangle$ and $|1\rangle$ with (respective) probabilities $|\alpha|^2$ and $|\beta|^2$.

7. For any density matrix ρ, the most natural mixture that yields ρ is its spectral decomposition: $\rho = \sum_i p_i |u_i\rangle\langle u_i|$, with eigenvectors $|u_i\rangle$ and eigenvalues p_i.

8. If the states $|u_1\rangle, \ldots, |u_m\rangle$ form an orthonormal basis, and each occurs with equal probability $1/m$, then the resulting density matrix is proportional to the identity:

$$\frac{1}{m}\sum_{i=1}^{m} |\psi_i\rangle\langle\psi_i| = \frac{1}{m}\mathbf{1}.$$

   This is called the **maximally mixed state**. For qubits, any pair of orthogonal states taken with equal probabilities gives the maximally mixed state $\frac{1}{2}\mathbf{1}$. In maximally mixed states, outcomes of *any* measurement are completely random.

9. It is often convenient to write density operators in terms of projectors on states which are not normalised, incorporating the probabilities into the length of the state vector:

$$\rho = \sum_i |\widetilde{\psi}_i\rangle\langle\widetilde{\psi}_i|$$

   where $|\widetilde{\psi}_i\rangle = \sqrt{p_i}|\psi_i\rangle$, i.e. $p_i = \langle\widetilde{\psi}_i|\widetilde{\psi}_i\rangle$. This form is more compact, but you have to remember that the state vectors are *not* normalised. We tend to mark such states

with the tilde, e.g. $|\widetilde{\psi}\rangle$, but you may have your own way to remember.

6.4 Mixed states of a qubit, and the Bloch ball

We have already talked in some depth about the Bloch sphere in Chapter 2 and Chapter 3, but now that we are considering density operators (which are strictly more general than state vectors), we are actually interested in the Bloch *ball*, i.e. not just the sphere of vectors of magnitude 1, but instead the ball of vectors of magnitude *less than or equal to* 1.

The most general Hermitian (2×2) matrix has four real parameters and can be expanded in the basis composed of the identity and the three Pauli matrices: $\{\mathbf{1}, \sigma_x, \sigma_y, \sigma_z\}$. Since the Pauli matrices are traceless, the coefficient of $\mathbf{1}$ in the expansion of a density matrix ρ must be $\frac{1}{2}$, so that $\operatorname{tr}\rho = 1$. Thus ρ may be expressed as

Physicists usually still refer to the Bloch *ball* as the Bloch *sphere*, even though it really is a ball now, not a sphere.

$$\begin{aligned}\rho &= \frac{1}{2}\left(\mathbf{1} + \vec{s}\cdot\vec{\sigma}\right) \\ &= \frac{1}{2}\left(\mathbf{1} + s_x\sigma_x + s_y\sigma_y + s_z\sigma_z\right) \\ &= \frac{1}{2}\begin{bmatrix} 1+s_z & s_x - is_y \\ s_x + is_y & 1-s_z \end{bmatrix}.\end{aligned}$$

The vector $\vec{s}$ is called the **Bloch vector** for the density operator ρ. Any real Bloch vector $\vec{s}$ defines a trace one Hermitian operator ρ, but in order for ρ to be a density operator it must also be non-negative. Which Bloch vectors yield legitimate density operators?

Let us compute the eigenvalues of ρ. The sum of the two eigenvalues of ρ is, of course, equal to one ($\operatorname{tr}\rho = 1$) and the product is equal to the determinant of ρ, which can be computed from the matrix form above:

$$\det\rho = \frac{1}{4}(1-s^2) = \frac{1}{2}(1+s)\frac{1}{2}(1-s)$$

where $s = |\vec{s}|$. It follows that the two eigenvalues of ρ are $\frac{1}{2}(1 \pm s)$. They have to be non-negative, and so s, the length of the Bloch vector, cannot exceed one. We can now visualise the convex set of (2×2) density matrices as a unit ball in three-dimensional Euclidean space: the extremal points, which represent pure states, are the points on the boundary ($s = 1$), i.e. the surface of the ball; the maximally mixed state $\mathbf{1}/2$ corresponds to $s = 0$, i.e. the centre of the ball. In general, the length of the Bloch vector s can be thought of as a "purity" of a state.

One might hope that there is an equally nice visualisation of the density operators in higher dimensions. Unfortunately there isn't.

6.5 Subsystems of entangled systems

We have already trumpeted that one of the most important features of the density operator formalism is its ability to describe the quantum state of a subsystem of a composite system. Let me now show you how it works.

Given a quantum state of the composite system $\mathcal{AB}$, described by some density operator $\rho^{\mathcal{AB}}$, we obtain reduced density operators $\rho^{\mathcal{A}}$ and $\rho^{\mathcal{B}}$ of subsystems $\mathcal{A}$ and $\mathcal{B}$, respectively, by the partial trace:

$$\begin{aligned}\rho^{\mathcal{AB}} &\longmapsto \underbrace{\rho^{\mathcal{A}} = \operatorname{tr}_{\mathcal{B}}\rho^{\mathcal{AB}}}_{\text{partial trace over }\mathcal{B}} \\ \rho^{\mathcal{AB}} &\longmapsto \underbrace{\rho^{\mathcal{B}} = \operatorname{tr}_{\mathcal{A}}\rho^{\mathcal{AB}}}_{\text{partial trace over }\mathcal{A}}\end{aligned}$$

We define the partial trace over $\mathcal{B}$, or $\mathcal{A}$, first on a tensor product of two operators $A \otimes B$ as

$$\begin{aligned}\mathrm{tr}_{\mathcal{B}}(A \otimes B) &= A(\mathrm{tr}\, B)\\ \mathrm{tr}_{\mathcal{A}}(A \otimes B) &= (\mathrm{tr}\, A)B,\end{aligned}$$

and then extend to any operator on $\mathcal{H}_{\mathcal{A}} \otimes \mathcal{H}_{\mathcal{B}}$ by linearity.

Here is a simple example. Suppose a composite system $\mathcal{AB}$ is in a pure entangled state, which we can always write as

$$|\psi_{\mathcal{AB}}\rangle = \sum_i c_i |a_i\rangle \otimes |b_i\rangle,$$

where $|a_i\rangle$ and $|b_j\rangle$ are two orthonormal bases (e.g. the Schmidt bases), and where $\sum_i |c_i|^2 = 1$ (due to the normalisation). The corresponding density operator of the composite system is the projector $\rho^{\mathcal{AB}} = |\psi_{\mathcal{AB}}\rangle\langle\psi_{\mathcal{AB}}|$, which we can write as

$$\rho^{\mathcal{AB}} = |\psi_{\mathcal{AB}}\rangle\langle\psi_{\mathcal{AB}}| = \sum_{ij} c_i c_j^\star |a_i\rangle\langle a_j| \otimes |b_i\rangle\langle b_j|$$

Let us compute the reduced density operator $\rho^{\mathcal{A}}$ by taking the partial trace over $\mathcal{B}$:

$$\begin{aligned}\rho^{\mathcal{A}} &= \mathrm{tr}_{\mathcal{B}}\, \rho^{\mathcal{AB}}\\ &= \mathrm{tr}_{\mathcal{B}}\, |\psi_{\mathcal{AB}}\rangle\langle\psi_{\mathcal{AB}}|\\ &= \mathrm{tr}_{\mathcal{B}} \sum_{ij} c_i c_j^\star |a_i\rangle\langle a_j| \otimes |b_i\rangle\langle b_j|\\ &= \sum_{ij} c_i c_j^\star |a_i\rangle\langle a_j| (\mathrm{tr}\, |b_i\rangle\langle b_j|)\\ &= \sum_{ij} c_i c_j^\star |a_i\rangle\langle a_j| \underbrace{\langle b_i|b_j\rangle}_{\delta_{ij}}\\ &= \sum_i |c_i|^2 |a_i\rangle\langle a_i|\end{aligned}$$

where we have used the fact that $\mathrm{tr}\, |b_i\rangle\langle b_j| = \langle b_j|b_i\rangle = \delta_{ij}$. In the $|a_i\rangle$ basis, the reduced density matrix $\rho^{\mathcal{A}}$ is diagonal, with entries $p_i = |c_i|^2$. We can also take the partial trace over $\mathcal{A}$ and obtain $\rho^{\mathcal{B}} = \sum_i |c_i|^2 |b_i\rangle\langle b_i|$. In particular, for the maximally entangled states in the $(d \times d)$-dimensional Hilbert spaces $\mathcal{H}_{\mathcal{A}} \otimes \mathcal{H}_{\mathcal{B}}$,

$$|\psi_{\mathcal{AB}}\rangle = \frac{1}{\sqrt{d}} \sum_i^d |a_i\rangle |b_i\rangle,$$

and the reduced density operators, $\rho^{\mathcal{A}}$ and $\rho^{\mathcal{B}}$, are the maximally mixed states: $\rho^{\mathcal{A}} = \rho^{\mathcal{B}} = \frac{1}{d}\mathbf{1}$. It follows that the quantum states of individual qubits in any of the Bell states are maximally mixed: their density matrix is $\frac{1}{2}\mathbf{1}$. A state such as

$$\frac{1}{\sqrt{2}} (|00\rangle + |11\rangle)$$

guarantees perfect correlations when each qubit is measured in the standard basis: the two equally likely outcomes are (0 and 0) or (1 and 1), but any single qubit outcome, be it 0 or 1 or anything else, is completely random.

6.6 Partial trace, revisited

If you are given a matrix you calculate the trace by summing its diagonal entries. How about the partial trace? Suppose someone writes down for you a density matrix of two qubits in the standard basis, $\{|00\rangle, |01\rangle, |10\rangle, |11\rangle\}$, and asks you to find the reduced density matrices of the individual qubits. The tensor product structure of this (4×4) matrix means that it is has a block form:

$$\rho^{\mathcal{AB}} = \left[\begin{array}{c|c} P & Q \\ \hline R & S \end{array}\right]$$

where P, Q, R, S are (2×2) sized sub-matrices. The two partial traces can then be evaluated as

$$\rho^{\mathcal{A}} = \mathrm{tr}_B\, \rho^{\mathcal{AB}} = \left[\begin{array}{c|c} \mathrm{tr}\, P & \mathrm{tr}\, Q \\ \hline \mathrm{tr}\, R & \mathrm{tr}\, S \end{array}\right]$$
$$\rho^{\mathcal{B}} = \mathrm{tr}_A\, \rho^{\mathcal{AB}} = P + S.$$

Take any of the Bell states, write its (4×4)-density matrix explicitly, and then trace over each qubit. In each case you should get the maximally mixed state.

The same holds for a general $\rho^{\mathcal{AB}}$ on any $\mathcal{H}_{\mathcal{A}} \otimes \mathcal{H}_{\mathcal{B}}$ with corresponding block form ($(m\times m)$ blocks of $(n \times n)$-sized sub-matrices, where m and n are the dimensions of $\mathcal{H}_{\mathcal{A}}$ and $\mathcal{H}_{\mathcal{B}}$, respectively).

6.7 Mixtures and subsystems

We have used the density operators to describe two distinct situations: the statistical properties of the mixtures of states, and the statistical properties of subsystems of composite systems. In order to see the relationship between the two, consider a joint state of a bipartite system $\mathcal{AB}$, written in a product basis in $\mathcal{H}_{\mathcal{A}} \otimes \mathcal{H}_{\mathcal{B}}$ as

$$\begin{aligned} |\psi_{\mathcal{AB}}\rangle &= \sum_{ij} c_{ij} |a_i\rangle \otimes |b_j\rangle \\ &= \sum_{j=1} |\widetilde{\psi}_j\rangle |b_j\rangle \\ &= \sum_{j=1} \sqrt{p_j} |\psi_j\rangle |b_j\rangle \end{aligned} \tag{9.7.1}$$

where $|\widetilde{\psi}_j\rangle = \sum_i c_{ij}|a_i\rangle = \sqrt{p_j}|\psi_j\rangle$, and the vectors $|\psi_j\rangle$ are the normalised versions of the $|\widetilde{\psi}_j\rangle$. Note that $p_j = \langle\widetilde{\psi}_j|\widetilde{\psi}_j\rangle$.

Then the partial trace over $\mathcal{B}$ gives the reduced density operator of subsystem $\mathcal{A}$:

$$\begin{aligned} \rho^{\mathcal{A}} &= \mathrm{tr}_{\mathcal{B}} \sum_{ij} |\widetilde{\psi}_i\rangle\langle\widetilde{\psi}_j| \otimes |b_i\rangle\langle b_j| \\ &= \sum_{ij} |\widetilde{\psi}_i\rangle\langle\widetilde{\psi}_j| (\mathrm{tr}\, |b_i\rangle\langle b_j|) \\ &= \sum_{ij} |\widetilde{\psi}_i\rangle\langle\widetilde{\psi}_j| \langle b_j|b_i\rangle \\ &= \sum_{i} |\widetilde{\psi}_i\rangle\langle\widetilde{\psi}_i| = \sum_i p_i |\psi_i\rangle\langle\psi_i|. \end{aligned}$$

Now, let us see how $\rho^{\mathcal{A}}$ can be understood in terms of mixtures. Let us place subsystems $\mathcal{A}$ and $\mathcal{B}$ in separate labs, run by Alice and Bob, respectively. When Bob measures part $\mathcal{B}$

in the $|b_j\rangle$ basis and obtains result k, which happens with the probability p_k, he prepares subsystem $\mathcal{A}$ in the state $|\psi_k\rangle$:

$$\sum_{i=1} \sqrt{p_j}|\psi_i\rangle|b_i\rangle \stackrel{\text{outcome } k}{\longmapsto} |\psi_k\rangle|b_k\rangle.$$

Bob does not communicate the outcome of his measurement. Thus, from Alice's perspective, Bob prepares a mixture of $|\psi_1\rangle, \ldots, |\psi_m\rangle$, with probabilities $p_1, \ldots, p_m$, which means that Alice, who knows the joint state but not the outcomes of Bob's measurement, may associate density matrix $\rho^{\mathcal{A}} = \sum_i p_i |\psi_i\rangle\langle\psi_i|$ with her subsystem $\mathcal{A}$. This is the same $\rho^{\mathcal{A}}$ which we obtained by the partial trace.

But suppose Bob chooses to measure his subsystem in some other basis. Will it have any impact on Alice's statistical predictions? Measurement in the new basis will result in a different mixture, but Alice's density operator will not change. Suppose Bob chooses basis $|d_i\rangle$ for his measurement. Any two orthonormal bases are connected by some unitary transformation, and so we can write $|b_i\rangle = U|d_i\rangle$ for some unitary U. In terms of components, $|b_i\rangle = \sum_j U_{ij}|d_j\rangle$. The joint state can now be expressed as

$$\begin{aligned}
|\psi_{\mathcal{AB}}\rangle &= \sum_i |\widetilde{\psi}_i\rangle|b_i\rangle \\
&= \sum_i |\widetilde{\psi}_i\rangle \left(\sum_j U_{ij}|d_j\rangle \right) \\
&= \sum_j \underbrace{\left(\sum_i U_{ij}|\widetilde{\psi}_i\rangle \right)}_{|\widetilde{\phi}_j\rangle} |d_j\rangle \\
&= \sum_j |\widetilde{\phi}_j\rangle|d_j\rangle.
\end{aligned}$$

If Bob measures in the $|d_i\rangle$ basis then he generates a new mixture of states $|\phi_1\rangle, \ldots |\phi_m\rangle$, which are the normalised versions of $|\widetilde{\phi}_1\rangle, \ldots |\widetilde{\phi}_m\rangle$, with each $|\phi_k\rangle$ occurring with probability $p_k = \langle\widetilde{\phi}_k|\widetilde{\phi}_k\rangle$. But this new mixture has exactly the same density operator as the previous one:

The U_{ij} are the components of a unitary matrix, hence $\sum_k U_{ik}U^{\star}_{jk} = \delta_{ij}$.

$$\begin{aligned}
\sum_j |\widetilde{\phi}_j\rangle\langle\widetilde{\phi}_j| &= \sum_{ijl} U_{ij}|\widetilde{\psi}_i\rangle\langle\widetilde{\psi}_l|U^{\star}_{lj} \\
&= \sum_{il} \underbrace{\left(\sum_j U_{ij}U^{\star}_{lj} \right)}_{\delta_{il}} |\widetilde{\psi}_i\rangle\langle\widetilde{\psi}_l| \\
&= \sum_i |\widetilde{\psi}_j\rangle\langle\widetilde{\psi}_j|,
\end{aligned}$$

which is exactly $\rho^{\mathcal{A}}$. So does it really matter whether Bob performs the measurement or not?

It does not.

After all, Alice and Bob may be miles away from each other, and if any of Bob's actions were to result in something that is physically detectable at the Alice's location that would amount to instantaneous communication between the two of them.

From the operational point of view it does not really matter whether the density opera-

tor represents our ignorance of the actual state (mixtures) or provides the only description we can have after discarding one part of an entangled state (partial trace). In the former case, the system is in some definite pure state but we do not know which. In contrast, when the density operator arises from tracing out irrelevant, or unavailable, degrees of freedom, the individual system cannot be thought to be in some definite state of which we are ignorant. Philosophy aside, the fact that the two interpretations give exactly the same predictions is useful: switching back and forth between the two pictures often offers additional insights and may even simplify lengthy calculations.

The two interpretations of density operators filled volumes of academic papers. The terms **proper mixtures** and **improper mixtures** are used, mostly by philosophers, to describe the statistical mixture and the partial trace approach, respectively.

6.8 Partial trace, yet again

The partial trace is the only map $\rho^{\mathcal{AB}} \mapsto \rho^{\mathcal{A}}$ such that

$$\mathrm{tr}[X\rho^{\mathcal{A}}] = \mathrm{tr}[(X \otimes \mathbf{1})\rho^{\mathcal{AB}}]$$

holds for any observable X acting on $\mathcal{A}$. This condition concerns the consistency of statistical predictions. Any observable X on $\mathcal{A}$ can be viewed as an observable $X \otimes \mathbf{1}$ on the composite system $\mathcal{AB}$, where $\mathbf{1}$ is the identity operator acting on $\mathcal{B}$. When constructing $\rho^{\mathcal{A}}$ we had better make sure that for any observable X the average value of X in the state $\rho^{\mathcal{A}}$ is the same as the average value of $X \otimes \mathbf{1}$ in the state $\rho^{\mathcal{AB}}$. This is indeed the case for the partial trace.

One can repeat the same argument for $\rho^{\mathcal{AB}} \mapsto \rho^{\mathcal{B}}$: the partial trace is the unique map $\rho^{\mathcal{AB}} \mapsto \rho^{\mathcal{B}}$ such that $\rho^{\mathcal{B}}$ satisfies $\mathrm{tr}[Y\rho^{\mathcal{B}}] = \mathrm{tr}[(1 \otimes Y)\rho^{\mathcal{AB}}]$ for any observable Y on $\mathcal{B}$.

For example, let us go back to the state in Equation (9.7.1) and assume that Alice measures some observable X on her part of the system. Technically, such an observable can be expressed as $X \otimes \mathbf{1}$, where $\mathbf{1}$ is the identity operator acting on the subsystem $\mathcal{B}$. The expected value of this observable in the state $|\psi_{\mathcal{AB}}\rangle$ is $\mathrm{tr}(X \otimes \mathbf{1})|\psi_{\mathcal{AB}}\rangle\langle\psi_{\mathcal{AB}}|$, i.e.

$$\begin{aligned}
\mathrm{tr}[(X \otimes \mathbf{1})\rho^{\mathcal{AB}}] &= \mathrm{tr}\left[(X \otimes \mathbf{1})\left(\sum_{ij} |\widetilde{\psi}_i\rangle\langle\widetilde{\psi}_j| \otimes |b_i\rangle\langle b_j|\right)\right] \\
&= \sum_{ij} \left[\mathrm{tr}\left(X|\widetilde{\psi}_i\rangle\langle\widetilde{\psi}_j|\right)\right] \underbrace{\left[\mathrm{tr}\left(|b_i\rangle\langle b_j|\right)\right]}_{\delta_{ij}} \\
&= \sum_i \mathrm{tr}\left[X|\widetilde{\psi}_i\rangle\langle\widetilde{\psi}_i|\right] \\
&= \mathrm{tr}\left[X \underbrace{\sum_i p_i|\psi_i\rangle\langle\psi_i|}_{\rho^{\mathcal{A}} = \mathrm{tr}_{\mathcal{B}}\, \rho^{\mathcal{AB}}}\right] \\
&= \mathrm{tr}[X\rho^{\mathcal{A}}]
\end{aligned}$$

as required.

!!!TODO!!! The uniqueness of the partial trace, for now see Nielsen & Chuang Box 2.6.

6.9 Remarks and exercises

6.9.1

Show that an arbitrary mixed state ρ can be represented as the partial trace $\mathrm{tr}\,|\psi\rangle\langle\psi|$ of a pure state of a larger system. Such a $|\psi\rangle$ is called a **purification** of ρ.

6.9.2

Show that purification is unique up to unitary equivalence. Let $|\psi_1\rangle$ and $|\psi_2\rangle$ in $\mathcal{H}_\mathcal{A} \otimes \mathcal{H}_\mathcal{B}$ be two pure states such that $\mathrm{tr}_\mathcal{B} |\psi_1\rangle\langle\psi_1| = \mathrm{tr}_\mathcal{B} |\psi_2\rangle\langle\psi_2|$. Show that $|\psi_1\rangle = \mathbf{1} \otimes U|\psi_2\rangle$ for some unitary operator U on $\mathcal{H}_\mathcal{B}$.

6.9.3

Two qubits are in the state described by the density operator $\rho = \rho^\mathcal{A} \otimes \rho^\mathcal{B}$. What is the partial trace of ρ over each qubit?

6.9.4

Write the density matrix of two qubits corresponding to the mixture of the Bell state $\frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$ with probability $\frac{1}{2}$ and the maximally mixed state of two qubits ((4×4) matrix in $\mathcal{H}_\mathcal{A} \otimes \mathcal{H}_\mathcal{B}$) with probability $\frac{1}{2}$.

6.9.5

!!!TODO!!! Trace norm

6.9.6

!!!TODO!!! How to distinguish between two different density operators

The power of interference and entanglement

PART

II

7 Bell's theorem

*About **quantum correlations**, which are stronger than any correlations allowed by classical physics, and about the **CHSH inequality** which demonstrates this fact, and which is a variant of **Bell's theorem**.*

Every now and then, it is nice to put down your lecture notes and go and see how things actually work in the real world. What is wonderful (and surprising) about quantum theory is that it turns up in many places that we might not expect it to, such as in the polarisation of light, where we stumble across an intriguing paradox.

If we take two polarising filters, and place them on top of each other with their polarisations oriented at 90° to one another, then basically no light will pass through, since the only light passing through the first filter is orthogonally polarised with respect to the second filter, and is thus blocked. But then, if we take a third filter, and place it in between the other two, at an angle in the middle of both (i.e. at 45°), then somehow *more* light is let through than if the middle filter weren't there at all.

For the more visually inclined, there is a video on YouTube by minutephysics about this experiment, or you can play with a virtual version at Quantum Flytrap.

This is intrinsically linked to Bell's theorem, which proves the technical sounding statement that "any local real hidden variable theory must satisfy certain statistical properties", which is *not* satisfied in reality, as many quantum mechanical experiments (such as the above) show!

7.1 Quantum correlations

Consider two entangled qubits in the singlet state

$$|\psi\rangle = \frac{1}{\sqrt{2}}\left(|01\rangle - |10\rangle\right)$$

and note that the projector $|\psi\rangle\langle\psi|$ can be written as

$$|\psi\rangle\langle\psi| = \frac{1}{4}\left(\mathbf{1}\otimes\mathbf{1} - \sigma_x\otimes\sigma_x - \sigma_y\otimes\sigma_y - \sigma_z\otimes\sigma_z\right).$$

There are other, more elementary, ways of deriving this result but here I want you to hone your skills. Now that you've learned about projectors, traces, and Pauli operators, why not put them to good use.

Any single qubit observable with values ± 1 can be represented by the operator

$$\vec{a}\cdot\vec{\sigma} = a_x\sigma_x + a_y\sigma_y + a_z\sigma_z,$$

where $\vec{a}$ is a unit vector in the three-dimensional Euclidean space. Suppose Alice and Bob choose measurements defined by vectors $\vec{a}$ and $\vec{b}$, respectively. For example, if the two qubits are spin-half particles, they may measure the spin components along the directions $\vec{a}$ and $\vec{b}$. We write the corresponding observable as the tensor product

$$A\otimes B = (\vec{a}\cdot\vec{\sigma})\otimes(\vec{b}\cdot\vec{\sigma}).$$

The eigenvalues of $A\otimes B$ are the products of eigenvalues of A and B. Thus $A\otimes B$ has two eigenvalues: $+1$, corresponding to the instances when Alice and Bob registered identical outcomes, i.e. $(+1,+1)$ or $(-1,-1)$; and -1, corresponding to the instances when Alice and Bob registered different outcomes, i.e. $(+1,-1)$ or $(-1,-1)$. This means that the expected value of $A\otimes B$, in any state, has a simple interpretation:

$$\langle A\otimes B\rangle = \Pr(\text{outcomes are the same}) - \Pr(\text{outcomes are different}).$$

This expression can take any numerical value from -1 (perfect anti-correlations) through 0 (no correlations) to $+1$ (perfect correlations). We now evaluate the expectation value in the singlet state:

$$\begin{aligned}\langle\psi|A\otimes B|\psi\rangle &= \mathrm{tr}\left[(\vec{a}\cdot\vec{\sigma})\otimes(\vec{b}\cdot\vec{\sigma})|\psi\rangle\langle\psi|\right]\\ &= -\frac{1}{4}\,\mathrm{tr}\left[(\vec{a}\cdot\vec{\sigma})\sigma_x\otimes(\vec{a}\cdot\vec{\sigma})\sigma_x+(\vec{a}\cdot\vec{\sigma})\sigma_y\otimes(\vec{a}\cdot\vec{\sigma})\sigma_y+(\vec{a}\cdot\vec{\sigma})\sigma_z\otimes(\vec{a}\cdot\vec{\sigma})\sigma_z\right]\\ &= -\frac{1}{4}\,\mathrm{tr}\left[(a_xb_x+a_yb_y+a_zb_z)\mathbf{1}\otimes\mathbf{1}\right]\\ &= -\vec{a}\cdot\vec{b}\end{aligned}$$

where we have used the fact that $\mathrm{tr}(\vec{a}\cdot\vec{\sigma})\sigma_k = a_k$ ($k=x,y,z$). So if Alice and Bob choose the same observable, $\vec{a}=\vec{b}$, then their outcomes will be always opposite: whenever Alice registers $+1$ (resp. -1) Bob is bound to register -1 (resp. $+1$).

7.2 Hidden variables

The story of "hidden variables" dates back to 1935 and grew out of Einstein's worries about the completeness of quantum theory. Consider, for example, a qubit. No quantum state of a qubit can be an eigenstate of two non-commuting operators, say σ_x and σ_z. If the qubit has a definite value of σ_x its value of σ_z must be indeterminate, and vice versa. If we take quantum theory to be a complete description of the world, then we must accept that it is impossible for both σ_x and σ_z to have definite values for the same qubit at the same time. Einstein felt very uncomfortable about all this. He argued that quantum theory is incomplete, and that observables σ_x and σ_z may both have simultaneous definite values, although we only have knowledge of one of them at a time. This is the hypothesis of **hidden variables**. In this view, the indeterminacy found in quantum theory is merely due to our ignorance of these "hidden variables" that are present in nature but not accounted for in the theory. Einstein came up with a number of pretty good arguments for the existence of "hidden variables". Probably the most compelling one was described in his 1935 paper (known as the EPR paper), co-authored with his younger colleagues, Boris Podolsky and Nathan Rosen. It stood for almost three decades as the most significant challenge to the completeness of quantum theory. Then, in 1964, John Bell showed that the hidden variable hypothesis can be tested and refuted.

7.3 CHSH inequality

*An upper bound on **classical** correlations.*

We will describe the most popular version of Bell's argument, introduced in 1969 by John Clauser, Michael Horne, Abner Shimony, and Richard Holt (CHSH). Let us assume that the results of any measurement on any individual system are predetermined. Any probabilities we may use to describe the system merely reflect our ignorance of these hidden variables.

Now, imagine the following scenario. Alice and Bob, two characters with a predilection for wacky experiments, are equipped with appropriate measuring devices and sent to two distant locations. Somewhere in between them there is a source that emits pairs of qubits that fly apart, one towards Alice and one towards Bob. Let us label the two qubits in each pair as $\mathcal{A}$ and $\mathcal{B}$ respectively, and let us assume that both Alice and Bob have well defined values of their observables. We ask Alice and Bob to measure one of the two pre-agreed observables. For each incoming qubit, Alice and Bob choose randomly, and independently

from each other, which particular observable will be measured. Alice chooses between A_1 and A_2, and Bob between B_1 and B_2. Each observable has value $+1$ or -1, and so we are allowed to think about them as random variables A_k and B_k, for $k = 1, 2$, which take values ± 1. Let us define a new random variable, the **CHSH quantity** S, as

$$S = A_1(B_1 - B_2) + A_2(B_1 + B_2).$$

It is easy to see that one of the terms $B_1 \pm B_2$ must be equal to zero and the other to ± 2, hence $S = \pm 2$. The average value of S must lie somewhere in-between, i.e.

$$-2 < \langle S \rangle < 2.$$

That's it! Such a simple and yet profound mathematical statement about correlations, which we refer simply to as the **CHSH inequality**. No quantum theory is involved because the CHSH inequality is not specific to quantum theory: it does not really matter what kind of physical process is behind the appearance of binary values of A_1, A_2, B_1, and B_2; it is a statement about correlations, and for all classical correlations we must have

$$|\langle A_1 B_1 \rangle - \langle A_1 B_2 \rangle + \langle A_2 B_1 \rangle + \langle A_2 B_2 \rangle| < 2.$$

There are essentially two two assumptions here:

1. **Hidden variables.** Observables have definite values
2. **Locality.** Alice's choice of measurements (A_1 or A_2) does not affect the outcomes of Bob's measurement, and vice versa.

We will not discuss the locality assumption here in detail but let me comment on it briefly. In the hidden variable world a, statement such as "if Bob were to measure B_1 then he would register $+1$" must be either true or false *prior to Bob's measurement*. Without the locality hypothesis, such a statement is ambiguous, since the value of B_1 could depend on whether A_1 or A_2 will be chosen by Alice. We do not want this for it implies the instantaneous communication. It means that, say, Alice by making a choice between A_1 and A_2, affects Bob's results. Bob can immediately "see" what Alice "does".

7.4 Quantum correlations revisited

In quantum theory the observables A_1, A_2, B_1, B_2 become 2×2 Hermitian matrices with two eigenvalues ± 1, and $\langle S \rangle$ becomes the expected value of the (4×4) **CHSH matrix**

$$S = A_1 \otimes (B_1 - B_2) + A_2 \otimes (B_1 + B_2).$$

We can now evaluate $\langle S \rangle$ using quantum theory. For example, if the two qubits are in the singlet state, then we know that

$$\langle A \otimes B \rangle = -\vec{a} \cdot \vec{b}.$$

We choose vectors $\vec{a}_1$, $\vec{a}_2$, $\vec{b}_1$, and $\vec{b}_2$ as shown in Figure 30, and then, with these choices,

$$\langle A_1 \otimes B_1 \rangle = \langle A_2 \otimes B_1 \rangle = \langle A_2 \otimes B_2 \rangle = \frac{1}{\sqrt{2}}$$
$$\langle A_1 \otimes B_2 \rangle = -\frac{1}{\sqrt{2}}.$$

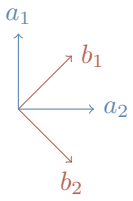


Figure 30. The relative angle between the two perpendicular pairs is $45°$.

Thus

$$\langle A_1 B_1\rangle - \langle A_1 B_2\rangle + \langle A_2 B_1\rangle + \langle A_2 B_2\rangle = -2\sqrt{2},$$

which obviously violates CHSH inequality.

To be clear, this violation has been observed in a number of painstakingly careful experiments. The early efforts were truly heroic, and the experiments had many layers of complexity. Today, however, such experiments are routine.

> The behaviour of entangled quantum systems cannot be explained by any local hidden variables.

7.5 Tsirelson's inequality

*An upper bound on **quantum** correlations.*

One may ask if $|\langle S\rangle| = 2\sqrt{2}$ is the maximal violation of the CHSH inequality, and the answer is "yes, it is": quantum correlations cannot achieve any value of $|\langle S\rangle|$ larger than $2\sqrt{2}$. This is because, for any state $|\psi\rangle$, the expected value $\langle S\rangle = \langle\psi|S|\psi\rangle$ cannot exceed the largest eigenvalue of S, and we can put an upper bound on the largest eigenvalues of S. To start with, the largest eigenvalue (in absolute value) of a Hermitian matrix M, denoted by $\|M\|$, is a matrix norm, and it has the following properties:

$$\begin{aligned}\|M\otimes N\| &= \|M\|\|N\|\\ \|MN\| &\leqslant \|M\|\|N\|\\ \|M+N\| &\leqslant \|M\| + \|N\|\end{aligned}$$

Given that $\|A_i\| = 1$ and $\|B_j\| = 1$ $(i,j = 1,2)$, it is easy to show that $\|S\| < 4$. One can, however, derive a tighter bound. We can show (do it) that

$$S^2 = 4(\mathbf{1}\otimes\mathbf{1}) + [A_1, A_2]\otimes[B_1, B_2].$$

The norm of each of the commutators ($\|[A_1, A_2]\|$ and $\|[B_1, B_2]\|$) cannot exceed 2, and $\|S^2\| = \|S\|^2$, which all together gives

$$\|S\| < 2\sqrt{2} \implies |\langle S\rangle| < 2\sqrt{2}.$$

This result is known as the **Tsirelson inequality**.

7.6 Remarks and Exercises

!!!TODO!!!

8 Quantum algorithms

About quantum interference in disguise: ***Hadamard, function evaluation, Hadamard****. Also about the early quantum algorithms and how they deal with querying oracles, searching for a needle in a haystack, and estimating periodicity of certain functions.*

Classical computers essentially evaluate functions: given n-bits of input they produce m-bits of output that are uniquely determined by the input; that is, they find the value of

$$f\colon \{0,1\}^n \to \{0,1\}^m$$

for a particular specified n-bit argument. A function with an m-bit value is equivalent to m Boolean functions, each with a one-bit value, so we may just as well say that the basic task performed by a computer is the evaluation of Boolean functions

$$f\colon \{0,1\}^n \to \{0,1\}.$$

How can we adapt this to the world of quantum computing?

8.1 Quantum Boolean function evaluation

In quantum computation, all elementary operations are reversible (unitary), so we compute Boolean functions in a reversible fashion as

$$|x\rangle|y\rangle \mapsto |x\rangle|y \oplus f(x)\rangle.$$

The corresponding circuit diagram (for $n = 3$) is shown in Figure 31.

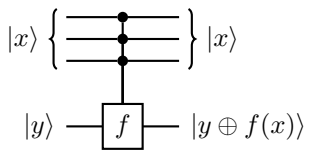


Figure 31. Computing some $f\colon \{0,1\}^3 \to \{0,1\}$ in a quantum manner.

Here we use two registers: the first one (counting from the top to the bottom of the circuit diagram) stores the arguments x, and the second one the values $f(x)$. More precisely, the value $f(x)$ is added bit-wise to the pre-existing binary value y of the second register. We usually set $y = 0$ to get

$$|x\rangle|0\rangle \mapsto |x\rangle|f(x)\rangle.$$

Quantum Boolean function evaluation is a special case of the generalised x-controlled-U on two registers:

$$\sum_x |x\rangle\langle x| \otimes U_x$$

where U_x is either the identity $\mathbf{1}$ (when $f(x) = 0$) or the bit-flip X (when $f(x) = 1$). We may also write this as

$$\sum_x |x\rangle\langle x| \otimes X^{f(x)}.$$

Do not confuse the capital X, which is the Pauli flip operator σ_x, with the small x, which is a binary string stored in the first register and the argument of our Boolean function f.

8.1.1 Example

Consider the Boolean function $f\colon \{0,1\}^2 \to \{0,1\}$ given by

$$f(x) = \begin{cases} 1 & \text{if } x = 01, \\ 0 & \text{otherwise.} \end{cases}$$

The evaluation $|x\rangle|y\rangle \mapsto |x\rangle|y \oplus f(x)\rangle$ can be tabulated explicitly as

$$\begin{array}{llll} |00\rangle|0\rangle \mapsto |00\rangle|0\rangle & |00\rangle|1\rangle \mapsto |00\rangle|1\rangle \\ |01\rangle|0\rangle \mapsto |01\rangle|1\rangle & |01\rangle|1\rangle \mapsto |01\rangle|0\rangle \\ |10\rangle|0\rangle \mapsto |10\rangle|0\rangle & |10\rangle|1\rangle \mapsto |10\rangle|1\rangle \\ |11\rangle|0\rangle \mapsto |11\rangle|0\rangle & |11\rangle|1\rangle \mapsto |11\rangle|1\rangle \end{array}$$

and the expression $\sum_x |x\rangle\langle x| \otimes X^{f(x)}$ becomes

$$\begin{aligned} &|00\rangle\langle 00| \otimes \mathbf{1} \\ +&|01\rangle\langle 01| \otimes X \\ +&|10\rangle\langle 10| \otimes \mathbf{1} \\ +&|11\rangle\langle 11| \otimes \mathbf{1}. \end{aligned}$$

Finally, the matrix form looks like

$$\left[\begin{array}{cc|cc|cc|cc} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{array}\right]$$

As you can see, this is a diagonal block matrix: a (4×4) matrix with (2×2) matrices as entries. The rows and the columns of the (4×4) matrix are labelled by the binary strings $00, 01, 10, 11$, and the (2×2) matrices on the diagonal represent operations applied to the qubit in the second register. Here all of them are the identity $\mathbf{1}$ except the $(01, 01)$ entry, which represents the bit-flip X. This is because $f(01) = 1$, and $f(x) = 0$ for all other binary strings x.

8.2 More phase kick-back

What makes the quantum evaluation of Boolean functions really interesting is its action on a superposition of different inputs x. For example,

$$\sum_x |x\rangle|0\rangle \longmapsto \sum_x |x\rangle|f(x)\rangle$$

produces $f(x)$ for *all* x in a *single* run (note that we have dropped the normalisation factor). It is more instructive to see the effect of the function evaluation when the qubit in the

second register is prepared in the state $|-\rangle = \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle)$, since then

$$\sum_x |x\rangle|-\rangle \longmapsto \sum_x (-1)^{f(x)}|x\rangle|-\rangle$$

(as shown in Figure 32). Whenever $f(x) = 1$, the bit flip X is applied to the qubit in the second register.

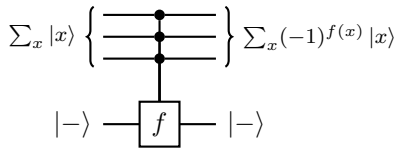


Figure 32. Computing some $f\colon \{0,1\}^3 \to \{0,1\}$ with the second-register qubit in state $|-\rangle$.

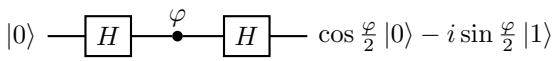
The reason for defining the state $|-\rangle$ as we do is that it is the eigenstate of X with eigenvalue -1. So, due the phase kick-back, whenever $f(x) = 1$, the phase factor -1 appears in front of the corresponding term $|x\rangle$. As you can see, the second register stays in state $|-\rangle$ all the way through the computation — it is the first register where things happen. Let us now see how quantum Boolean function evaluation introduces phase shifts in quantum interference experiments, and how such experiments can be viewed as computations.

8.3 Oracles and query complexity

The computational power of quantum interference was discovered by counting how many times certain Boolean functions have to be evaluated in order to find the answer to a given problem. Imagine a "black box" (also called an **oracle**) that computes some fixed Boolean function, but whose inner workings are unknown to us, and a scenario in which one wants to learn about a given property of the Boolean function but has to "pay" (in energy, or in money!) for each use (often referred to as a **query**) of the box. In such a setting, the objective is to minimise number of queries to the oracle while finding out as much information as possible about the function computed by the oracle. For this purpose, we ignore everything that happens inside the black box: the Boolean function evaluation counts as just *one* computational step.

8.3.1 Deutsch's algorithm

We start, once more, with the simplest quantum interference circuit:



Suppose you can prepare the input, you can read the output, you *cannot* see the phase shifter, *but* you are promised that the phase shifter is set to either $\varphi = 0$ or $\varphi = \pi$. Can you tell which value φ has been set to?

Of course you can!

One way of doing it is to set your input to $|0\rangle$ and check the output: for $\varphi = 0$ the output is always $|0\rangle$, and for $\varphi = \pi$ it is always $|1\rangle$. A single run of the interference experiment is sufficient to determine the difference. The first quantum algorithm, proposed by David Deutsch in 1985, is very much related to this effect, but where the phase setting is determined by the Boolean function evaluation via the phase kick-back.

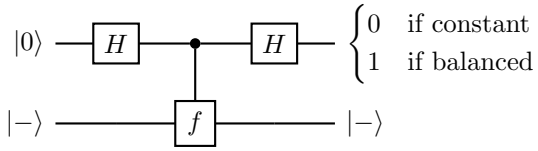
Scenario. We are presented with an oracle that computes some unknown function $f\colon \{0,1\} \to \{0,1\}$. Note that there are only four possibilities for what f can be: it could be one of two constant functions (i.e. those where $f(0) = f(1)$), or one of two "balanced" functions (i.e. those where $f(0) \neq f(1)$).

	$f(0)$	$f(1)$
constant	0	0
	1	1
balanced	0	1
	1	0

Our task is to determine, using the fewest queries possible, whether the function computed by the oracle is constant or balanced.

Note that we are *not* asked for the particular values $f(0)$ and $f(1)$, but *only whether the two values are the same or different*. Classical intuition tells us that we have to evaluate both $f(0)$ and $f(1)$ and compare them, which involves evaluating f *twice*. But, in the quantum setting, we can solve this problem with a *single* function evaluation, using the following circuit.

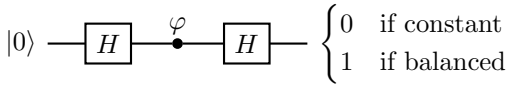
Circuit. (Deutsch's).
First register: 1 *qubit. Second register:* 1 *qubit.*



During the function evaluation, the second register "kicks back" the phase factor $(-1)^{f(x)}$ in front of $|x\rangle$, but the state of the second register remains unchanged; the first register is modified as follows:

$$
\begin{aligned}
|0\rangle &\xmapsto{H} |0\rangle + |1\rangle \\
&\xmapsto{f} (-1)^{f(0)}|0\rangle + (-1)^{f(1)}|1\rangle \\
&\equiv |0\rangle + (-1)^{f(0)\oplus f(1)}|1\rangle \\
&\xmapsto{H} |f(0) \oplus f(1)\rangle.
\end{aligned}
$$

This evolution can be represented by the circuit diagram



where the relative phase is $\varphi = (-1)^{f(0)\oplus f(1)}$. The first qubit ends in state $|0\rangle$ if the function f is constant, and in state $|1\rangle$ if the function is balanced, and the standard measurement distinguishes these two cases with certainty.

The original Deutsch algorithm provides the correct answer with probability 50%. Here we have presented a modified/improved version.

Deutsch's result laid the foundation for the new field of quantum computation, and was followed by several other quantum algorithms for various problems. They all seem to rest on the same generic sequence: a Hadamard transform, followed by a function evaluation, followed by another Hadamard (or Fourier) transform. As we shall see in a moment, in some cases (such as in Grover's search algorithm) this sequence is repeated several times. Let me now take you through the three early quantum algorithms, each one offering a higher-order speed-up when compared to their classical analogues than the last.

The Hadamard transform is a special case of the Fourier transform over the group $\mathbb{Z}_2^n$.

8.4 Three more quantum algorithms

Along with Deutsch's algorithm, there are three more fundamental quantum algorithms that we will study here. Each one was designed to solve a different specific problem, but they all share some similarity: this omnipresent sequence of Hadamard, function evaluation, Hadamard.

8.4.1 The Bernstein-Vazirani algorithm

Scenario. We are presented with an oracle that computes some unknown function $f\colon \{0,1\}^n \to \{0,1\}$, but we are promised that f is of the form

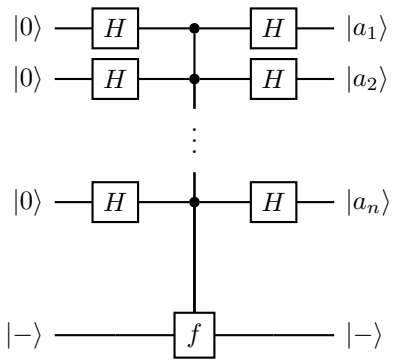
$$f(x) = a \cdot x \equiv (a_1 \cdot x_1) \oplus \ldots \oplus (a_n \cdot x_n)$$

for some fixed $a \in \{0,1\}^n$.
Our task is to determine, using the fewest queries possible, the value of the n-bit string a.

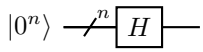
It's quite easy to see how to do this classically: if we input the value $x = 00\ldots010\ldots0$, with the 1 in the m-th bit, then $f(x)$ is simply the m-th bit of a; after n such calls, we can evaluate every bit value. It is also clear that there cannot exist a better classical algorithm: each call to the oracle teaches us exactly one bit of information, and since we must learn n bits, we must query it n times.

In contrast, by running the circuit below, it is possible to determine the value of a with a *single* (!) call to the oracle.

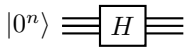
Circuit. (Bernstein-Vazirani).
First register: n *qubits. Second register:* 1 *qubit.*



N.B. The "…" in the circuit means "there are more wires here but they are identical (apart from the numbering) to the ones above". You might also see other notation to denote this, such as



or even simply



Stepping through the execution of the circuit (and ignoring the second register, which remains in state $|-\rangle$ throughout), we obtain

$$
\begin{aligned}
|0\rangle &\xmapsto{H} \left(\frac{1}{\sqrt{2}}\right)^n \sum_{x\in\{0,1\}^n} |x\rangle \\
&\xmapsto{f} \left(\frac{1}{\sqrt{2}}\right)^n \sum_{x\in\{0,1\}^n} (-1)^{a\cdot x}|x\rangle \\
&\xmapsto{H} \left(\frac{1}{\sqrt{2}}\right)^n \sum_{x\in\{0,1\}^n} \left[(-1)^{a\cdot x} \left(\frac{1}{\sqrt{2}}\right)^n \sum_{y\in\{0,1\}^n} (-1)^{y\cdot x}|y\rangle\right] \\
&= \left(\frac{1}{2}\right)^n \sum_{y\in\{0,1\}^n} \left[\sum_{x\in\{0,1\}^n} (-1)^{(a\oplus y)\cdot x}\right] |y\rangle \\
&= |a\rangle
\end{aligned}
$$

where we write the second Hadamard transform as

$$|x\rangle \mapsto \left(\frac{1}{\sqrt{2}}\right)^n \sum_{y\in\{0,1\}^n} (-1)^{y\cdot x}|y\rangle$$

and where we have used the fact (which you should prove!) that, for any $y \in \{0,1\}^n$,

$$\sum_{x\in\{0,1\}^n} (-1)^{x\cdot y} = \begin{cases} 0 & \text{if } y \neq 0 \\ 2^n & \text{if } y = 0 \end{cases}$$

This lets us write

$$\sum_{x\in\{0,1\}^n} (-1)^{(a\oplus y)\cdot x} = \begin{cases} 0 & \text{if } y \neq a \\ 2^n & \text{if } y = a. \end{cases}$$

Even if you don't immediately see how this sum works for $z \neq a$ (writing $|z\rangle$ to mean the output), you can first calculate the probability that the output is $z = a$. In this case it is easy to see that the sum is 2^n, and that in the final state $\sum_z \alpha_z|z\rangle$ the term $z = a$ has amplitude 1. Thus, by normalisation, all the other terms must be equal to 0.

If you take the sum over x, then all the terms always cancel out *unless* $a \oplus y = 00\ldots0$, i.e. *unless* $y = a$. Thus the standard bit-by-bit measurement of the first register gives the value of a and solves the problem with a single call to the oracle.

Note that the algorithm follows the same pattern as Deutsch's algorithm: Hadamard, function evaluation, Hadamard, i.e. a generic quantum interference pattern.

8.4.2 Grover's search algorithm

The next algorithm we will study aims to solve the problem of searching for a specific item in an *unsorted* database. Think about an old-fashioned phone book: the entries are

typically sorted alphabetically, by the name of the person that you want to find. However, what if you were in the opposite situation: you had a phone number and wanted to find the corresponding person's name? The phone book is not sorted in that way, and to find the number (and hence name) with, say, 50% probability, you would need to search through, on average, 50% of the entries. In a large phone book, this would take a long time.

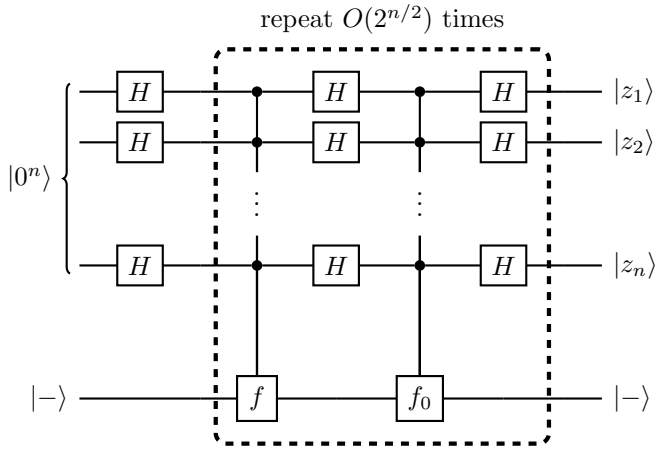
While this might seem like a rather contrived problem (a computer database should always maintain an index on any searchable field), many problems in computer science can be cast in this form, i.e. that of an **unstructured search**.

Scenario. We are presented with an oracle that computes some unknown function $f\colon \{0,1\}^n \to \{0,1\}$.
Our task is to find, using the fewest queries possible, an input $x \in \{0,1\}^n$ such that $f(x) = 1$.

Suppose that we know that, amongst the $N = 2^n$ binary strings, there are $M \ll N$ which are "tagged", i.e. on which f evaluates to 1. There is no structure in the database, and so any classical search requires around N/M steps, i.e. the function f must be evaluated roughly N/M times.

In contrast, there is a quantum search algorithm, implemented by the circuit below, that was proposed in 1996 by Lov Grover, and which requires only roughly $\sqrt{N/M}$ steps.

Circuit. (Grover's search).
First register: n qubits. Second register: 1 qubit.



where f_0 tags the binary string of n zeros: $f_0(x) = 1$ if $x = 00\ldots0$, and $f_0(x) = 0$ otherwise.

We can recognise the typical Hadamard, function evaluation, Hadamard sequence, and we can see that the second register (the bottom qubit, in state $|-\rangle$) plays an auxiliary role: the real action takes place in the first register. However, unlike the previous algorithms, a single call to the oracle does not do very much, and we have to build up the quantum interference in the first register through repeated calls to the oracle (without any intermediate measurements!).

Here, the basic step is the **Grover iteration operator** G, which is the boxed part of the circuit that we repeat over and over. After $O(2^{n/2})$ applications of G, we measure the first

register bit-by-bit and obtain the value of $|z\rangle$, which is such that, *with "high" probability*, $f(z) = 1$. In order to actually *see* how this algorithm works, and to justify our claim that it gives what we are searching for "with high probability", we can take a more geometric approach.

First, we define two orthonormal vectors in the Hilbert space of the first register:

In fact, we shall completely ignore the second register from now on.

$$|a\rangle = \frac{1}{\sqrt{N-M}} \sum_{x \in f^{-1}(0)} |x\rangle$$
$$|b\rangle = \frac{1}{\sqrt{M}} \sum_{x \in f^{-1}(1)} |x\rangle$$

where $f^{-1}(i) = \{x \in \{0,1\}^n \mid f(x) = i\}$. These two vectors span a two-dimensional subspace in which the search will take place.

This subspace contains the equally-weighted superposition $|s\rangle$ of all binary strings of length n:

We often omit from our notation the fact that the sum is over all $x \in \{0,1\}^n$, leaving it (hopefully) implicitly understood from the context.

$$\begin{aligned} |s\rangle &= \frac{1}{\sqrt{N}} \sum_x |x\rangle \\ &= \sqrt{\frac{N-M}{N}}|a\rangle + \sqrt{\frac{M}{N}}|b\rangle \\ &= (\cos\alpha)|a\rangle + (\sin\alpha)|b\rangle \end{aligned}$$

where we have parametrised $\sqrt{\frac{N-M}{N}}$ as $\cos\alpha$, and $\sqrt{\frac{M}{N}}$ as $\sin\alpha$, with $\alpha \approx \sqrt{\frac{M}{N}}$, since $N \gg M$.

The state $|s\rangle$ is our starting input for our sequence of Grover iterations, and we will show that, applying G, when restricting to the plane spanned by $|a\rangle$ and $|b\rangle$, amounts to applying a rotation by angle 2α. Grover's search algorithm can then be understood as a sequence of rotations which take the input state $|s\rangle$ towards the target state $|b\rangle$.

To see this, note that the oracle induces the unitary transformation

$$f\colon |x\rangle \mapsto (-1)^{f(x)}|x\rangle$$

which we shall write as $I_a = 2|a\rangle\langle a| - \mathbf{1}$, and interpret as a reflection through the $|a\rangle$-axis. In particular, evaluation of f_0 can be written as $2|0\rangle\langle 0| - \mathbf{1}$, and thus thought of as a reflection through the $|0\rangle$-axis. If we sandwich f_0 in between two Hadamards then we obtain $I_s = 2|s\rangle\langle s| - \mathbf{1}$, which is reflection through the $|s\rangle$-axis. The Grover iteration operator G is the composition

$$G = I_s I_a.$$

Note also that $I_a = 2|a\rangle\langle a| - \mathbf{1} = \mathbf{1} - 2|b\rangle\langle b|$.

Now recall the purely geometric fact that (working in 2-dimensional Euclidean space), if we have two intersecting lines L_1 and L_2, meeting with angle θ, then reflecting an object through L_1 and then reflecting the resulting image through L_2 is the same as simply rotating the original object around the point of intersection $L_1 \cap L_2$ by 2θ.

The angle between $|a\rangle$ and $|s\rangle$ is α, and so, each time G is applied, the vector is rotated (around the origin) by 2α towards the $|b\rangle$-axis. We just have to choose the right number r of steps such that we end up as close to the $|b\rangle$-axis as possible. The state $|s\rangle$ starts at angle α to $|a\rangle$, and we should perform our final (and only) measurement when this angle is $\pi/2$,

i.e. when $(2r+1)\alpha = \pi/2$, which gives

$$r \approx \frac{\pi}{4}\sqrt{\frac{N}{M}}.$$

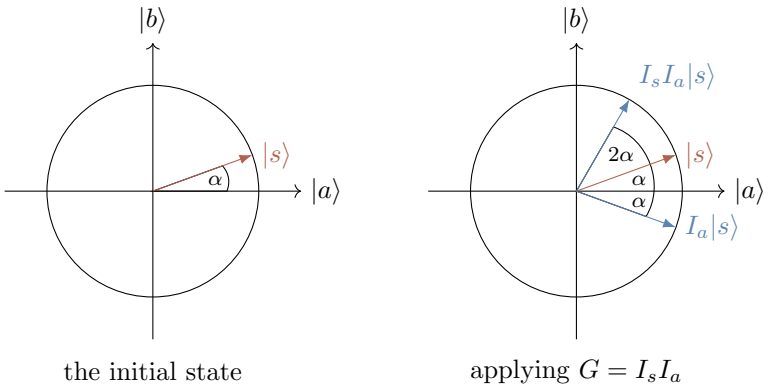


Figure 33. Understanding the Grover search algorithm geometrically.

So the quantum algorithm searches an unsorted list of N items in roughly $\sqrt{N}$ steps: it offers a *quadratic* speed-up when compared to classical search, which can be of immense practical importance. For example, cracking some of the popular ciphers, such as AES (Advanced Encryption Standard), essentially requires a search among *many* binary strings (called **keys**). If these can be checked at a rate of, say, one million keys per second, then a classical computer would need over a thousand years to find the correct key, while a quantum computer using Grover's algorithm would find it in less than four minutes.

8.4.3 Simon's problem

Here we will see the simplest quantum algorithm that shows an exponential speed-up compared to the best classical algorithm.

Scenario. We are presented with an oracle that computes some unknown function $f\colon \{0,1\}^n \to \{0,1\}^n$, but we are promised that f satisfies, for all $x \in \{0,1\}^n$,

$$f(x) = f(x \oplus s)$$

for some fixed $s \in \{0,1\}^n$, which we call the **period** of f. So that the problem is non-trivial (i.e. so that f really is two-to-one), we assume that s is *not* the string of n zeros.
Our task is to determine, using the fewest queries possible, the value of the n-bit string s.

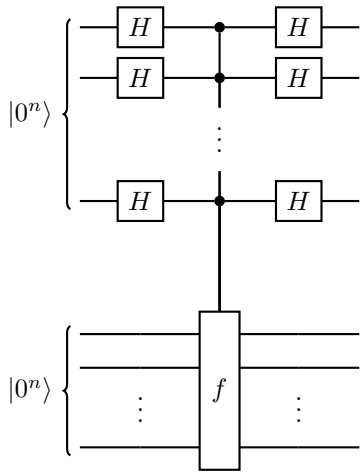
This is equivalent to saying that f is **two-to-one**: for any $y \in \{0,1\}^n$ such that there exists some $x \in \{0,1\}^n$ with $f(x) = y$, there exists exactly one other $x' \neq x$ such that $f(x') = y$ as well.

Classically, this problem is exponentially hard. We will not go through a detailed proof of this fact, but the intuition is reasonably simple: since there is no structure in the function f that would help us find its period s, the best we can do is evaluate f on random inputs and hope that we find some distinct x and x' such that $f(x) = f(x')$, and then we know that $s = x \oplus y$. After having made m queries to the oracle, we have a list of m values of the tuple $(x, f(x))$; there are $m(m-1)/2$ possible pairs which could match within this list, and the probability that a randomly chosen pair match is $1/2^{n-1}$. This means that the

probability of there being at least one matching pair within the list of m tuples is less than $m^2/2^n$. Clearly, the chance of finding a matching pair is negligible if the oracle is queried on fewer than $\sqrt{2^n}$ inputs.

The quantum case, on the other hand, gives a result with high probability within a *linear* number of steps. The circuit that solves this problem, shown below, has a familiar Hadamard–function–Hadamard structure, but the second register has been expanded to n qubits.

Circuit. (Simon's problem).
First register: n qubits. Second register: n qubits.



Let's follow the evolution of the two registers in this circuit. We start off by preparing the equally-weighted superposition of all n-bit strings, and then query the oracle:

$$\begin{aligned} |0^n\rangle|0^n\rangle &\longmapsto \frac{1}{\sqrt{2^n}}\sum_x |x\rangle|0^n\rangle \\ &\longmapsto \frac{1}{\sqrt{2^n}}\sum_x |x\rangle|f(x)\rangle. \end{aligned}$$

The second Hadamard transform on the first register then yields the final output state:

$$\frac{1}{2^n}\sum_{x,y}(-1)^{x\cdot y}|y\rangle|f(x)\rangle. \tag{6.4.3.1}$$

Now, if we measure the second register *before* applying the second Hadamard transform to the first, we obtain one of the 2^{n-1} possible values of $f(x)$, each equally likely.

As we shall see in a moment, the actual measurement on the second register is not actually necessary.

Suppose that the outcome of the measurement is $f(a)$. Given that both a and $a \oplus s$ are mapped to $f(a)$ by f, the first register then collapses to the state

$$\frac{1}{\sqrt{2}}\big(|a\rangle + |a \oplus s\rangle\big).$$

The subsequent Hadamard transform on the first register then gives us the final state

We write $s^\perp$ to mean the set of all $y \in \{0,1\}^n$ such that $y \cdot s = 0$.

$$\frac{1}{\sqrt{2^{n+1}}} \sum_y (-1)^{a \cdot y} \Big(1 + (-1)^{s \cdot y}\Big) |y\rangle |f(a)\rangle = \frac{1}{\sqrt{2^{n-1}}} \sum_{y \in s^\perp} (-1)^{a \cdot y} |y\rangle |f(a)\rangle$$

where we have used the fact that $(a \oplus s) \cdot y = (a \cdot y) \oplus (s \cdot y)$, and that $1 + (-1)^{s \cdot y}$ can have only two values: either 2 (when $s \cdot y = 0$), or 0 (when $s \cdot y = 1$). Now we measure the first register: the outcome is selected at random from all possible values of y such that $a \cdot y = 0$, each occurring with probability $1/(2^{n-1})$.

In fact, we do not have to measure the second register at all: it was a mathematical shortcut, simply taken for pedagogical purposes. Instead of collapsing the state to just one term in a superposition, we can express Equation (6.4.3.1) as

Recall that the image of f is the set of all $z \in \{0,1\}^n$ such that there exists some $x \in \{0,1\}^n$ satisfying $f(x) = z$.

$$\frac{1}{2^n} \sum_{y, f(a)} \Big((-1)^{a \cdot y} + (-1)^{(a \oplus s) \cdot y}\Big) |y\rangle |f(a)\rangle = \frac{1}{2^n} \sum_{y, f(a)} (-1)^{a \cdot y} \Big(1 + (-1)^{s \cdot y}\Big) |y\rangle |f(a)\rangle$$

where the summation over $f(a)$ means summing over all binary strings in the image of f.

The output of the algorithm is then

$$\frac{1}{2^{n-1}} \sum_{y \in s^\perp} |y\rangle \sum_{f(a)} (-1)^{a \cdot y} |f(a)\rangle$$

and, again, the measurement outcome is selected at random from all possible values of y such that $s \cdot y = 0$.

We are not quite done yet: we cannot infer s from a *single* output y. However, once we have found $n-1$ linearly independent strings $y_1, y_2, \ldots, y_{n-1}$, we can solve the $n-1$ equations

Here, **linearly independent** means that no string in the set $\{y_1, \ldots, y_n\}$ can be expressed as the bitwise sum of some other strings in this set.

$$\begin{Bmatrix} s \cdot y_1 = 0 \\ s \cdot y_2 = 0 \\ \vdots \\ s \cdot y_{n-1} = 0 \end{Bmatrix}$$

to determine a unique value of s. (Note that we only need $n-1$ values, and not n, because $s = 0$ will always be a solution, but we have explicitly assumed that this is not the case, and so it suffices to narrow down the space of possible solutions to consist of *two* elements, since then we know that we can just take the non-zero one.)

So we run this algorithm repeatedly, each time obtaining another value of y that satisfies $s \cdot y = 0$. Every time we find some new value of y that is linearly independent of all previous ones, we can discard half the potential candidates for s. The probability that $y_1, \ldots, y_{n-1}$ are linearly independent is

$$\left(1 - \frac{1}{2^{n-1}}\right) \left(1 - \frac{1}{2^{n-2}}\right) \cdots \left(1 - \frac{1}{2}\right). \tag{6.4.3.2}$$

Indeed, suppose that we have k linearly independent binary strings $y_1, \ldots, y_k$. Then these strings span a subspace of size 2^k, consisting of all binary strings of the form $\bigoplus_{i=1}^k b_i y_i$, where $b_1, \ldots, b_k \in \{0,1\}$. Now suppose we obtain some y_{k+1}. It will be linearly independent from the $y_1, \ldots, y_k$ if and only if it lies *outside* the subspace spanned by the $y_1, \ldots, y_k$, which occurs with probability $1 - (2^k)/(2^n)$. We can bound Equation (6.4.3.2) from be-

low: the probability of obtaining a linearly independent set $\{y_1, \ldots, y_{n-1}\}$ by running the algorithm $n-1$ times (i.e. not having to discard any values and run again) is

$$\prod_{k=1}^{n-1}\left(1-\frac{1}{2^k}\right) \geqslant \left[1-\left(\frac{1}{2^{n-1}}+\frac{1}{2^{n-2}}+\ldots+\frac{1}{4}\right)\right]\cdot\frac{1}{2} > \frac{1}{4}.$$

Use the inequality

$$\begin{aligned}(1-x)(1-y) &= 1-x-y-xy \\ &\geqslant 1-(x+y)\end{aligned}$$

which holds for any $0 < x, y < 1$.

We conclude that we can determine s with some constant probability of error after repeating the algorithm $O(n)$ times. The exponential separation that this algorithm demonstrates between quantum and classical highlights the vast potential of a quantum computer to speed up function evaluation.

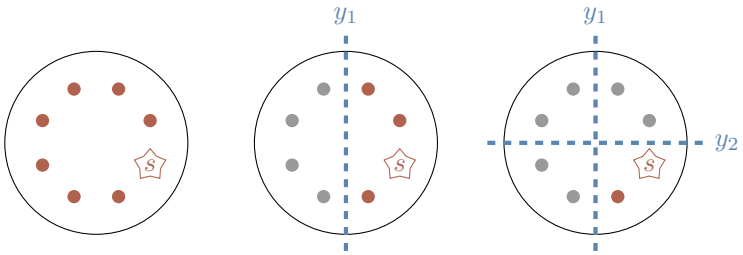


Figure 34. Picture all possible binary strings as dots, but with the string s denoted by a star. Every linearly independent y_{k+1} lets us "zoom in" twice as close towards s.

8.5 Remarks and exercises

8.5.1

Consider the Boolean function $f\colon \{0,1\}^n \to \{0,1\}$ defined by $f(x) = a \cdot x$ for some fixed $a \in \{0,1\}^n$. Exactly one half of the binary strings $x \in \{0,1\}^n$ give $f(x) = 0$, and the other half give $f(x) = 1$.

8.5.2

!!!TODO!!! implementing reflections

8.5.3

!!!TODO!!! optimality of Grover

9 Decoherence, and elements of quantum error correction

*About the one big problem that hinders us from physically implementing everything that we've learnt so far, namely **decoherence**, as well as how we can start to deal with it via some elementary **error correction**.*

In principle we know how to build a quantum computer: we can start with simple quantum logic gates and try to integrate them together into quantum networks. However, if we keep on putting quantum gates together into networks we will quickly run into some serious practical problems. The more interacting qubits involved, the harder it is to prevent them from getting entangled with the environment. This unwelcome entanglement, also known as **decoherence**, destroys the interference, and thus the power, of quantum computing.

9.1 Decoherence simplified

Consider the following qubit-environment interaction:

$$|0\rangle|e\rangle \longmapsto |0\rangle|e_{00}\rangle$$
$$|1\rangle|e\rangle \longmapsto |1\rangle|e_{11}\rangle$$

where $|e\rangle$, $|e_{00}\rangle$, and $|e_{11}\rangle$ are the states of the environment, which not need to be orthogonal. Let $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$ be the initial state of the qubit. The environment is essentially trying to *measure* the qubit and, as the result, the two get entangled:

The reason we use two indices in $|e_{00}\rangle$ and $|e_{11}\rangle$ will become clear in a moment, when we consider more general interaction with the environment.

$$\Big(\alpha|0\rangle + \beta|1\rangle\Big)|e\rangle \longmapsto \alpha|0\rangle|e_{00}\rangle + \beta|1\rangle|e_{11}\rangle.$$

This state can also be written as

$$\Big(\alpha|0\rangle + \beta|1\rangle\Big)|e\rangle \longmapsto \Big(\alpha|0\rangle + \beta|1\rangle\Big)\frac{|e_{00}\rangle + |e_{11}\rangle}{2} + \Big(\alpha|0\rangle - \beta|1\rangle\Big)\frac{|e_{00}\rangle - |e_{11}\rangle}{2}.$$

or as

$$|\psi\rangle|e\rangle \longmapsto \mathbf{1}|\psi\rangle|e_{\mathbf{1}}\rangle + Z|\psi\rangle|e_Z\rangle,$$

where $|e_{\mathbf{1}}\rangle = \frac{1}{2}(|e_{00}\rangle + |e_{11}\rangle)$ and $|e_Z\rangle = \frac{1}{2}(|e_{00}\rangle - |e_{11}\rangle)$. We may interpret this expression by saying that two things can happen to the qubit: nothing $\mathbf{1}$ (first term), or phase-flip Z (second term). *This, however, should not be taken literally unless the states of the environment, $|e_{\mathbf{1}}\rangle$ and $|e_Z\rangle$, are orthogonal.*

Why not?

This process is what we refer to as **decoherence**.

9.2 Decoherence and interference

Suppose the qubit undergoes the usual interference experiment, but, in between the two Hadamard gates, it is affected by **decoherence** (denoted by $\times$), which acts as described above (i.e. $|0\rangle|e\rangle \mapsto |0\rangle|e_{00}\rangle$ and $|1\rangle|e\rangle \mapsto |1\rangle|e_{11}\rangle$).

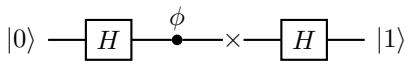


Figure 35. The usual interference experiment, but with decoherence.

Let us step through the circuit in Figure 35, keeping track of the state of the environment:

$$\begin{aligned}
|0\rangle|e\rangle &\overset{H}{\longmapsto} \Big(|0\rangle + |1\rangle\Big)|e\rangle \\
&\overset{\phi}{\longmapsto} \Big(|0\rangle + e^{i\phi}|1\rangle\Big)|e\rangle \\
&\overset{\times}{\longmapsto} |0\rangle|e_{00}\rangle + e^{i\phi}|1\rangle|e_{11}\rangle \\
&\overset{H}{\longmapsto} |0\rangle\Big(|e_{00}\rangle + e^{i\phi}|e_{11}\rangle\Big) + |1\rangle\Big(|e_{00}\rangle - e^{i\phi}|e_{11}\rangle\Big).
\end{aligned}$$

If we write $\langle e_{00}|e_{11}\rangle = ve^{i\alpha}$, then the final probabilities of 0 and 1 oscillate with ϕ as

$$\begin{aligned}
P_0(\phi) &= \frac{1}{2}\big(1 + v\cos(\phi+\alpha)\big), \\
P_1(\phi) &= \frac{1}{2}\big(1 - v\cos(\phi+\alpha)\big).
\end{aligned}$$

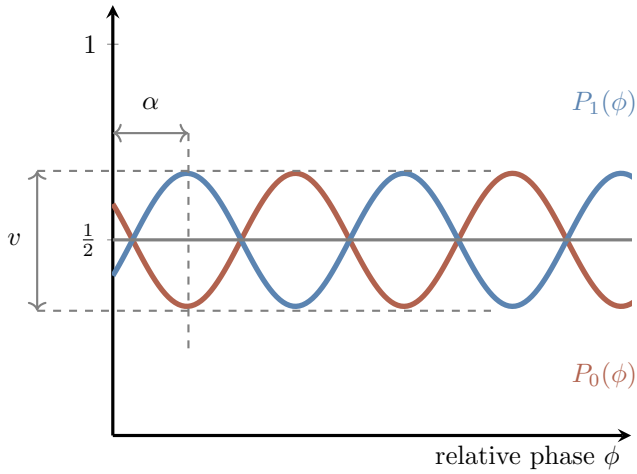


Figure 36. Visibility suppression.

As we can see in Figure 36, the interference pattern is suppressed by a factor v, which we call the **visibility**. As $v = |\langle e_{00}|e_{11}\rangle|$ decreases, we lose all the advantages of quantum interference. For example, in Deutsch's algorithm we obtain the correct answer with probability at most $\frac{1}{2}(1+v)$. For $\langle e_{00}|e_{11}\rangle = 0$, the **perfect decoherence** case, the network outputs 0 or 1 with equal probabilities, i.e. *it is useless as a computing device*.

> It is clear that we want to avoid decoherence, or at least diminish its impact on our computing device. For this we need **quantum error correction**: we encode the state of a single (logical) qubit across several (physical) qubits.

!!!TODO!!! generalised decoherence as controlled-U gate, varying from 1 to controlled-`NOT`

9.3 Evolution of density operators under decoherence

In terms of density operators, the qubit alone evolves from the pure state $|\psi\rangle\langle\psi|$ to a mixed state, which can be obtained by tracing over the environment. We start with the evolution

of the state vector $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$, which is given by

$$(\alpha|0\rangle + \beta|1\rangle)\,|e\rangle \longmapsto \alpha|0\rangle|e_{00}\rangle + \beta|1\rangle|e_{11}\rangle,$$

Then we write it as the evolution of the projector $|\psi\rangle\langle\psi|$, and trace over the environment to obtain

$$\begin{aligned}|\psi\rangle\langle\psi| \longmapsto &|\alpha|^2|0\rangle\langle 0|\langle e_{00}|e_{00}\rangle + \alpha\beta^\star|0\rangle\langle 1|\langle e_{11}|e_{00}\rangle \\ &+\alpha^\star\beta|1\rangle\langle 0|\langle e_{00}|e_{11}\rangle + |\beta|^2|1\rangle\langle 1|\langle e_{11}|e_{11}\rangle.\end{aligned}$$

Written in the matrix form, this is

$$\begin{bmatrix}|\alpha|^2 & \alpha\beta^* \\ \alpha^*\beta & |\beta|^2\end{bmatrix} \longmapsto \begin{bmatrix}|\alpha|^2 & \alpha\beta^*\langle e_{11}|e_{00}\rangle \\ \alpha^*\beta\langle e_{00}|e_{11}\rangle & |\beta|^2\end{bmatrix}.$$

The off-diagonal elements, originally called **coherences**, vanish as $\langle e_{00}|e_{11}\rangle$ approaches zero. This is why this particular interaction is called decoherence.

Notice that

$$|\psi\rangle|e\rangle \longmapsto \mathbf{1}|\psi\rangle|e_{\mathbf{1}}\rangle + Z|\psi\rangle|e_Z\rangle,$$

implies

$$|\psi\rangle\langle\psi| \longmapsto \mathbf{1}|\psi\rangle\langle\psi|\mathbf{1}\langle e_{\mathbf{1}}|e_{\mathbf{1}}\rangle + Z|\psi\rangle\langle\psi|Z\langle e_Z|e_Z\rangle,$$

only when $\langle e_{\mathbf{1}}|e_Z\rangle = 0$ (otherwise you would have additional cross terms $\mathbf{1}|\psi\rangle\langle\psi|Z$ and $Z|\psi\rangle\langle\psi|\mathbf{1}$). In this case we can indeed say that, with probability $\langle e_{\mathbf{1}}|e_{\mathbf{1}}\rangle$, nothing happens, and, with probability $\langle e_Z|e_Z\rangle$, the qubit undergoes the phase-flip Z.

9.4 Quantum errors

The most general qubit-environment interaction

$$\begin{aligned}|0\rangle|e\rangle &\longmapsto |0\rangle|e_{00}\rangle + |1\rangle|e_{01}\rangle, \\ |1\rangle|e\rangle &\longmapsto |1\rangle|e_{10}\rangle + |0\rangle|e_{11}\rangle,\end{aligned}$$

where the states of the environment are neither normalised nor orthogonal, leads to decoherence

$$\begin{aligned}\Big(\alpha|0\rangle + \beta|1\rangle\Big)|e\rangle \longmapsto &\Big(\alpha|0\rangle + \beta|1\rangle\Big)\frac{|e_{00}\rangle + |e_{11}\rangle}{2} \\ &+\Big(\alpha|0\rangle - \beta|1\rangle\Big)\frac{|e_{00}\rangle - |e_{11}\rangle}{2} \\ &+\Big(\alpha|1\rangle + \beta|0\rangle\Big)\frac{|e_{01}\rangle + |e_{10}\rangle}{2} \\ &+\Big(\alpha|1\rangle - \beta|0\rangle\Big)\frac{|e_{01}\rangle - |e_{10}\rangle}{2}.\end{aligned}$$

We can also write this as

$$|\psi\rangle|e\rangle \longmapsto \mathbf{1}|\psi\rangle|e_{\mathbf{1}}\rangle + Z|\psi\rangle|e_Z\rangle + X|\psi\rangle|e_X\rangle + Y|\psi\rangle|e_Y\rangle.$$

The intuition behind this expression is that four things can happen to the qubit:

1. nothing ($\mathbf{1}$),
2. phase-flip (Z),
3. bit-flip (X), or
4. both bit-flip and phase-flip (Y).

This is certainly the case when the states $|e_{\mathbf{1}}\rangle, |e_X\rangle, |e_Y\rangle$ and $|e_Z\rangle$ are mutually orthogonal, otherwise we cannot perfectly distinguish between the four alternatives.

What is important here is the discretisation of errors, and the fact that we can reduce quantum errors to *two types*: bit-flip errors X, and phase-flip errors Z.

In general, given n qubits in state $|\psi\rangle$ and the environment in state $|e\rangle$ the joint evolution can be expanded as

$$|\psi\rangle|e\rangle \longmapsto \sum_i E_i|\psi\rangle|e_i\rangle,$$

where the E_i are the n-fold tensor products of the Pauli operators and the $|e_i\rangle$ are the corresponding states of the environment, which are not assumed to be normalised or mutually orthogonal. A typical operator E_i acting on five qubits may look like this,

$$X \otimes Z \otimes \mathbf{1} \otimes \mathbf{1} \otimes Y \equiv XZ\mathbf{1}\mathbf{1}Y.$$

We can say that E_i represents an error consisting of the bit (X) error on the first qubit, phase (Z) error on the second qubit and both bit and phase (Y) error on the fifth qubit. Again, *this is not quite accurate if the corresponding states of the environment are not mutually orthogonal*, but it gives the right kind of intuition nonetheless. Here the index i in E_i ranges from 1 to $4^5 = 1024$, because there are 4^5 different Pauli operators acting on 5 qubits.

9.5 Same evolution, different errors

We can always pick up an orthonormal basis $|u_i\rangle$ in the environment and express the system–environment evolution as

$$\begin{aligned}|\psi\rangle|e\rangle &\longmapsto \sum_{ij} E_i|\psi\rangle|u_j\rangle\langle u_j|e_i\rangle \\ &= \sum_j \Big(\sum_i \langle u_j|e_i\rangle E_i\Big)|\psi\rangle|u_j\rangle \\ &= \sum_j M_j|\psi\rangle|u_j\rangle.\end{aligned}$$

The new "error" operators M_j satisfy $\sum_j M_j^\dagger M_j = \mathbf{1}$ and, in general, they are *not* unitary. Now, the evolution of the density operator $|\psi\rangle\langle\psi|$ can be written as

$$|\psi\rangle\langle\psi| \longmapsto \sum_j M_j|\psi\rangle\langle\psi|M_j^\dagger.$$

Which particular errors you choose depends of your choice of the basis in the environment. If, instead of $|u_j\rangle$, you pick up a different basis, say $|v_k\rangle$, then

$$\begin{aligned}|\psi\rangle|e\rangle \longmapsto & \sum_j M_j|\psi\rangle|u_j\rangle \\ &= \sum_j M_j|\psi\rangle \sum_k |v_k\rangle\langle v_k|u_j\rangle \\ &= \sum_k \Big(\sum_j \langle v_k|u_j\rangle M_j\Big)|\psi\rangle|v_k\rangle \\ &= \sum_k N_k|\psi\rangle|v_k\rangle,\end{aligned}$$

and, consequently,

$$|\psi\rangle\langle\psi| \longmapsto \sum_k N_k|\psi\rangle\langle\psi|N_k^\dagger.$$

The new "errors" satisfy $\sum_k N_k^\dagger N_k = \mathbf{1}$, and the error operators N_k and M_j are related by the unitary matrix $U_{kj} = \langle v_k|u_j\rangle$.

9.6 Some errors can be corrected on some states

Alice prepares a quantum object in some state $|\psi\rangle$ and sends it to Bob. The object is intercepted by a malicious Eve who changes its state by applying one of the prescribed unitary operations $U_1, \ldots, U_n$, with probabilities $p_1, \ldots, p_n$, respectively. Alice and Bob know the set of possible unitaries (errors), and the associated probabilities, but they do not know which particular unitary operation was chosen by Eve. Can Bob reconstruct the state $|\psi\rangle$? The answer is affirmative, at least for *some states* $|\psi\rangle$.

Let $\mathcal{H}$ be the Hilbert space pertaining to the object, and let $\mathcal{C}$ be a subspace of $\mathcal{H}$. Suppose $|\psi\rangle \in \mathcal{C}$, and that, for each vector in $\mathcal{C}$, we have

$$\langle\psi|U_i^\dagger U_j|\psi\rangle = \delta_{ij}$$

Any error U_k transforms the subspace $\mathcal{C}$ into the subspace $\mathcal{C}_k$, which is orthogonal to $\mathcal{C}$ and also to any other subspace $\mathcal{C}_j$ for $j \neq k$. All Bob has to do is - perform a measurement, defined by projectors on the subspaces $\mathcal{C}_j$ for $j = 1, \ldots n$, - identify k, and - apply $U_k^\dagger$.

As an example, consider an object composed of three qubits and the subspace $\mathcal{C}$ spanned by the two basis vectors $|000\rangle$ and $|111\rangle$. Suppose Eve applies one of the following four unitary operations: $U_0 = \mathbf{1}\otimes\mathbf{1}\otimes\mathbf{1}$, $U_1 = X\otimes\mathbf{1}\otimes\mathbf{1}$, $U_2 = \mathbf{1}\otimes X\otimes\mathbf{1}$, and $U_3 = \mathbf{1}\otimes\mathbf{1}\otimes X$. That is, the identity, or bit-flip on the first, second, or third qubit. Each operation is chosen randomly with the same probability of $1/4$. It is easy to see that the four operations generate four subspaces:

$$\begin{aligned}\mathcal{C} &= \Big\langle |000\rangle, |111\rangle \Big\rangle \qquad \mathcal{C}_1 = \Big\langle |100\rangle, |011\rangle \Big\rangle \\ \mathcal{C}_2 &= \Big\langle |010\rangle, |101\rangle \Big\rangle \qquad \mathcal{C}_3 = \Big\langle |001\rangle, |110\rangle \Big\rangle.\end{aligned}$$

The eight dimensional Hilbert space of the three qubits is then decomposed into the sum of orthogonal subspaces

$$\mathcal{C} \oplus \mathcal{C}_1 \oplus \mathcal{C}_2 \oplus \mathcal{C}_3$$

So suppose Alice prepares $|\psi\rangle = \alpha|000\rangle + \beta|111\rangle$ and Eve applies the bit-flip to the third qubit. This generates the state $\mathbf{1} \otimes \mathbf{1} \otimes X|\psi\rangle = \alpha|001\rangle + \beta|110\rangle \in \mathcal{C}_3$. The projective measurement on these subspaces tells Bob that the new state is in the subspace $\mathcal{C}_3$, and hence the original state can be recovered by the operation $\mathbf{1} \otimes \mathbf{1} \otimes X$.

9.7 Repetition codes

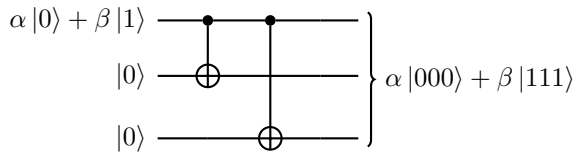
In order to give a sense of how quantum error correction actually works, let us begin with a *classical* example of a repetition code. Suppose a transmission channel flips each bit in transit with probability p. If this error rate is considered too high then it can be decreased by encoding each bit into, say, three bits:

$$0 \mapsto 000$$
$$1 \mapsto 111.$$

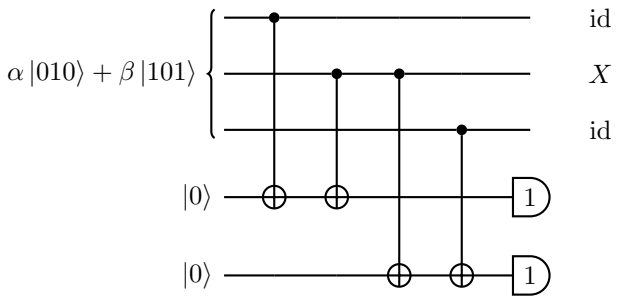
That is, each time we want to send logical 0, we send three physical bits, all in state 0; each time we want to send logical 1, we send three physical bits, all in state 1. The receiver decodes the bit value by a "majority vote" of the three bits. If only one error occurs, then this error correction procedure is foolproof. In general, the net probability of error is just the likelihood that two or three errors occur, which is $3p^2(1-p) + p^3 < p$. Thus the three bit code improves the reliability of the information transfer. The *quantum* case, however, is more complicated, because we have both bit-flip *and* phase-flip errors.

9.8 Quantum error correction

In order to protect a qubit against bit-flips (incoherent X rotations), we rely on the same repetition code, but both encoding and error correction is now done by quantum operations. We take a qubit in some unknown pure state $\alpha|0\rangle + \beta|1\rangle$, introduce two auxiliary qubits, and encode it into three qubits as



Suppose that at most one qubit is then flipped (say, the second one). The encoded state then becomes $\alpha|010\rangle + \beta|101\rangle$. Decoding requires some care: if we measure the three qubits directly it would destroy the superposition of states that we are working so hard to protect. Instead we introduce another two additional qubits, both in state $|0\rangle$, and apply the following network:



We measure the two auxiliary qubits, also known as **ancilla bits**, and the result of the measurement, known as the **error syndrome**, tells us how to reset the three qubits of the code. The theory behind this network runs as follows.

If qubits one and two (counting from the top) are the same, then the first ancilla is in the $|0\rangle$ state. Similarly, if qubits two and three are the same, then the second ancilla is in the $|0\rangle$ state. However, if they are different, then the corresponding ancilla is in the $|1\rangle$ state. Hence, the four possible error syndromes — $|00\rangle$, $|01\rangle$, $|10\rangle$, and $|11\rangle$ — each indicate a different possibility: no errors, an error in the third, first, or second qubits (respectively). In our example, we would measure $|11\rangle$, revealing that both qubits 1 and 2, and qubits 2 and 3, are different. Thus it is qubit 2 that has an error. Knowing the error, we can go back and fix it, simply by applying X to qubit 2. The net result is the state $\alpha|000\rangle + \beta|111\rangle$, which is then turned into $(\alpha|0\rangle + \beta|1\rangle)|0\rangle|0\rangle$ by running the mirror image of the encoding network.

9.9 Turning bit-flips into phase-flips

The three-qubit code that we have just demonstrated is sufficient to protect a qubit against single bit-flips, but not phase-flips. But this is good enough. Recall that $HZH = X$, and so it is enough to sandwich the decoherence area in between the Hadamard gates: they will turn phase flips into bit flips, and we already know hot to protect our qubits against Z-errors. The encoded state $\alpha|0\rangle+\beta|1\rangle$ now reads $\alpha|{+}{+}{+}\rangle+\beta|{-}{-}{-}\rangle$, where $|\pm\rangle = |0\rangle\pm|1\rangle$.

9.10 Dealing with bit-flip and phase-flip errors

We can now put the bit-flip and phase-flip codes together: first we encode the qubit using the phase-flip code, and then we encode each of the three qubits of the code using the bit-flip code. This gives an error correction scheme that allows us to protect against both types of error, thus yielding a code that encodes a single logical qubit across nine physical qubits, protecting against a single quantum error on any of the nine qubits.

If we want to preserve a quantum state for a long time without doing any computations, or if we want to send it through a noisy communications channel, we can just encode the state using a quantum code and decode it when we are done. Computation on encoded states using noisy gates requires few more tricks (to be completed).

9.11 Remarks and Exercises

!!!TODO!!!

Quantum security

PART

III

10 Quantum channels, or CP maps

*About the linear transformations — often called **quantum channels**, **quantum operations**, or **superoperators** — that map density operators to density operators. Mathematically, such objects are known as **completely positive trace-preserving maps**. Be prepared for some name dropping; you will hear about Karl Kraus, Woody Stinespring, Andrzej Jamiokowski, and Man-Duen Choi. To be sure, knowing names will not give you more insights, but at least you will not be intimidated when you hear about the **Stinespring** and the **Kraus representations**, the **Jamiokowski isomorphism**, or the **Choi matrix**.*

In quantum information science (and in life) we often divide the world into two parts: things we can control and everything else. Things we can control (at least with some precision), such as a bunch of ions in an ion trap, will now acquire a generic name — the **system** — and the "everything else" part will be called the **environment**. The joint evolution of the system and the environment is unitary, but the system *alone* may deviate from the unitary evolution due to the entangling interactions with the environment. What does this non-unitary quantum evolution look like?

Mathematically speaking, we want to characterise **physically admissible** maps

$$\mathcal{E}\colon \mathscr{B}(\mathcal{H}) \to \mathscr{B}(\mathcal{H}')$$

that map density operators to density operators. We will call them **quantum channels**. Here, $\mathcal{H}$ and $\mathcal{H}'$ are the Hilbert spaces associated with the input and the output, respectively (these two spaces may be of different dimensions: $\dim \mathcal{H} = d$ and $\dim \mathcal{H}' = d'$, with d not necessarily equal to d'). It is not a trivial task, since some operations (such as the matrix transpose) turn density matrices into seemingly legal density matrices, and yet they are *not* physically admissible. We want to understand why.

Recall that, given a pair of Hilbert spaces $\mathcal{H}_\mathcal{A}$ and $\mathcal{H}_\mathcal{B}$, we denote the set of (bounded) linear operators from $\mathcal{H}_\mathcal{A}$ to $\mathcal{H}_\mathcal{B}$ by $\mathscr{B}(\mathcal{H}_\mathcal{A}, \mathcal{H}_\mathcal{B})$. The shorthand $\mathscr{B}(\mathcal{H})$ is used to denote $\mathscr{B}(\mathcal{H}, \mathcal{H})$. Thus an element of $\mathscr{B}(\mathcal{H})$ is exactly a density operator on $\mathcal{H}$.

There are essentially two approaches to characterising quantum channels. Let us call them **constructive** and **axiomatic**.

- In the constructive approach (much favoured by physicists) we assume that, in the quantum world, there are only unitary evolutions, and that *non*-unitary evolutions of subsystems are induced by unitary evolutions of *larger* systems. We can view them as the result of adding or removing physical systems that participate in unitary evolutions. In particular, *any* quantum channel can be constructed by taking a physical system in state ρ, combining it with an auxiliary system (e.g. the environment) in some fixed quantum state, applying a unitary transformation to the combined system, and then discarding everything that is of no interest to us.

- In the axiomatic approach we proceed as mathematicians usually do. We refer to quantum theory for inspiration, scratch our heads, and list all the desired properties that a map should have in order to be called an admissible quantum operation. This way we define **completely positive trace-preserving** (CPTP) maps.

In the end, you will be relieved to learn, the two approaches are equivalent, leading to the same mathematical formalism. This equivalence is often goes by the name of the **Stinespring dilation theorem**, which says that *CPTP maps are exactly the maps that can be constructed from unitary evolutions by adding and removing subsystems*.

William Forrest "Woody" Stinespring (1929–2012) was an American mathematician specialising in operator theory.

10.1 The constructive approach

10.1.1 Unitary evolution is all there is...

Here we consider the constructive approach described above. So we know that any evolution of any *isolated* system is always unitary, since, at the fundamental level, this is the only evolution that is offered by quantum theory. Anything else that is not unitary has to be, somehow, derived from a unitary evolution of the whole, and it is easy to see that a unitary evolution of the whole does not imply unitary evolutions of its constituent parts. The question is: *what kind of sub-evolutions can be induced on subsystems by a global unitary evolution?*

Consider a physical system, prepared in some input state $|\psi\rangle \in \mathcal{H}$, which is then combined with another system (here, the environment) that is initially in state $|e\rangle \in \mathcal{H}_E$. The two systems interact, and their joint unitary evolution U produces, in general, an entangled state. The system itself, after discarding the environment, ends in an output state ρ', which is given by taking a partial trace, as shown in Figure 37.

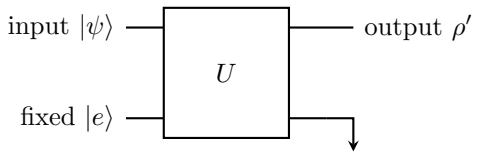


Figure 37. The output (discarding what happens to the environment) is given by $\rho' = \mathrm{tr}_E\left(U|\psi\rangle\langle\psi|\otimes |e\rangle\langle e|U^\dagger\right)$. The arrow at the end of the second quantum wire represents discarding that state.

The requirement that the environment start in a pure state is not very restrictive: if it were initially in a mixed state, then we could always regard the environment as a subenvironment of some larger environment in an entangled pure state. It *is* important, however, that the system and its environment *initially have no correlation with each other*: they are in a product state $|\psi\rangle \otimes |e\rangle$. Then, if we fix the initial state of the environment $|e\rangle$ and the unitary U, the final state of the system depends *only* on its initial state $|\psi\rangle$. In fact, the initial state of the system can be a mixed state, described by some density operator ρ. This is because any mixed state can be interpreted as a statistical mixture of pure states, and any linear process respects such a mixture. Thus, by fixing U and $|e\rangle$, we have constructed a well defined linear map $\mathcal{E}\colon \mathscr{B}(\mathcal{H}) \to \mathscr{B}(\mathcal{H})$ that transforms the input state ρ into the output state $\rho' = \mathcal{E}(\rho)$.

$$\begin{aligned} \mathcal{E}\colon \mathscr{B}(\mathcal{H}) &\longrightarrow \mathscr{B}(\mathcal{H}) \\ \rho &\longmapsto \rho' = \mathrm{tr}_E\left(U|\psi\rangle\langle\psi| \otimes |e\rangle\langle e|U^\dagger\right) \end{aligned}$$

This way of describing the action of a quantum channel is often called the **unitary representation** of a channel. As mentioned above, the channel is completely defined by the unitary U and the initial state of the environment $|e\rangle$, but note that, once the system and the environment cease to interact, any operation *on the environment alone* has no effect on the state of the system. That is, the two diagrams in Figure 38 define the same channel,

or, in other words, the unitaries U and $(\mathbf{1} \otimes R)U$, where R acts only on $\mathcal{H}_E$, define the same channel.

Consider U and $U(\mathbf{1}\otimes R)$, where R acts only on $\mathcal{H}_E$. Do these two unitaries also define the same quantum channel?

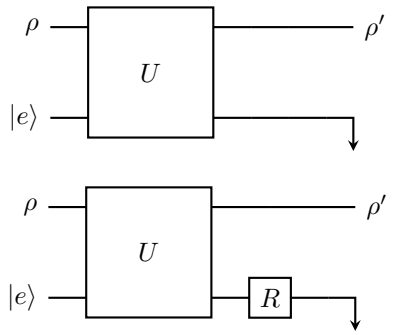


Figure 38. The two circuits here define the same quantum channel.

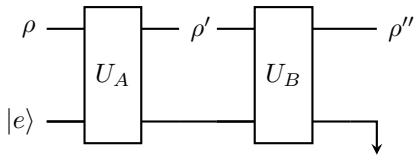
Note that, for any operator $\tilde{\rho}$ on $\mathcal{H} \otimes \mathcal{H}_E$, and for any operator R on $\mathcal{H}_E$, we have that

$$\mathrm{tr}_E \left[(\mathbf{1} \otimes R)\tilde{\rho}(\mathbf{1} \otimes R^\dagger)\right] = \mathrm{tr}_E\, \tilde{\rho}.$$

The reduced density operator $\rho = \mathrm{tr}_E\, \tilde{\rho}$ is not affected by R. We can easily prove this for operators $\tilde{\rho}$ that are tensor products $\tilde{\rho} = X \otimes Y$ (a good exercise to try yourself), and then, by linearity, extend the result to any operator $\tilde{\rho}$.

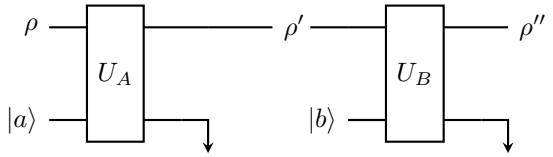
We must point out that our construction of quantum channels does not capture all possible quantum evolutions. The assumption that the system and the environment are not initially correlated is crucial, and it does impose some restrictions on the applicability of our formalism. Compare, for example, the following two versions of a process in which the system, initially in state ρ, undergoes two stages of evolution.

Firstly:



Here, the environment is not discarded after the first unitary evolution U_A — it carries on and participates in the second unitary evolution U_B. In this case, the evolutions $\rho \mapsto \rho'$ and $\rho \mapsto \rho''$ *are* well-defined quantum channels, but the evolution $\rho' \mapsto \rho''$ is *not*; it falls outside the remit of our formalism for the input state of the system and the state of the environment are not independent.

Secondly:



In contrast to the previous process, here we discard the environment after the first unitary, and start the second unitary evolution in a fresh tensor product state with a *new* environment — the two stages involve independent environments. In this case, all three

evolutions ($\rho \mapsto \rho'$, $\rho' \mapsto \rho''$, and $\rho \mapsto \rho''$) are well-defined quantum channels, and can be composed. If $\mathcal{E}_A$ describes the evolution from ρ to ρ', and $\mathcal{E}_B$ the evolution from ρ' to ρ'', then the composition $\mathcal{E}_B\mathcal{E}_A$ describes the evolution from ρ to ρ''.

Unitary evolutions form a group; quantum channels form a *semigroup*. Quantum channels are invertible only if they are unitary operations or simple isometric embeddings, such as including the environment in some fixed state and then discarding it right away, without any intermediate interaction.

In practice, we often deal with complex environments that have internal dynamics that "hide" any entanglement with the system as quickly as it arises. For example, let our system be an atom, surrounded by electromagnetic field (which serves as the environment), and let the field start in the vacuum state. If the atom emits a photon into the environment, then the photon quickly propagates away, and the immediate vicinity of the atom appears to be empty, i.e. it resets to the vacuum state. In this approximation, we assume that the environment quickly forgets about its state resulting from any previous evolution. This is known as the **Markov approximation**, and, in the quantum Markov process, the environment has essentially no memory.

Andrey Markov (1929–2012) was a Russian mathematician best known for his work on stochastic processes.

10.1.2 Kraus operators

The unitary representation of a quantum channel is usually too general to be of any practical use, at least for our purposes. We will find it convenient, as you will see when we start discussing quantum error correction, to choose an orthonormal basis $\{|e_k\rangle\}$ of $\mathcal{H}_E$, and express the joint (i.e. "system plus environment") unitary evolution U as

$$U(|\psi\rangle \otimes |e\rangle) = \sum_k E_k|\psi\rangle \otimes |e_k\rangle$$

where E_k is an operator on $\mathcal{H}$ defined by

$$E_k|\psi\rangle = \langle e_k|U(|\psi\rangle \otimes |e\rangle)$$

for any vector $|\psi\rangle \in \mathcal{H}$ Think about $\langle e_k|U(|\psi\rangle\otimes|e\rangle)$ as a "partial inner product" of $|e_k\rangle \in \mathcal{H}_E$ and $U|\psi\rangle|e\rangle \in \mathcal{H} \otimes \mathcal{H}_E$. In analogy with classical communication channels we sometimes refer to the operators E_k as **quantum errors**, but they are more generally referred to as **Kraus operators**.

Given $\mathcal{H}_\mathcal{A} \otimes \mathcal{H}_\mathcal{B}$, the **partial inner product** of $|x\rangle \in \mathcal{H}_\mathcal{A}$ and $|a\rangle \otimes |b\rangle \in \mathcal{H}_\mathcal{A}\otimes\mathcal{H}_\mathcal{B}$ is first defined by the formula

$$\langle x|(|a\rangle \otimes |b\rangle) = \langle x|a\rangle|b\rangle$$

and then extended to other (non-direct tensor product) vectors in $\mathcal{H}_\mathcal{A} \otimes \mathcal{H}_\mathcal{B}$ by linearity. We discuss this further in the exercises.

The intuition behind the expression

$$|\psi\rangle \otimes |e\rangle \longmapsto \sum_k E_k|\psi\rangle \otimes |e_k\rangle$$

is that the operation E_k (which changes the state of the system as $|\psi\rangle \mapsto E_k|\psi\rangle$) occurs with the probability $p_k = \langle\psi|E_k^\dagger E_k|\psi\rangle$. The probabilities sum up to one, i.e. $\sum_k p_k = \langle\psi|E_k^\dagger E_k|\psi\rangle = 1$ for any state $|\psi\rangle$, and so the operators E_k must satisfy the completeness relation

$$\sum_k E_k^\dagger E_k = \mathbf{1}.$$

Note that the matrix U acts on the tensor product space, which gives it a natural partition into blocks of sub-matrices, such as $E_k = \langle e_k|U|e\rangle$. In order to visualise $E_k = \langle e_k|U|e\rangle$, write U as a matrix on $\mathcal{H}_E \otimes \mathcal{H}$ rather than $\mathcal{H} \otimes \mathcal{H}_E$ (i.e. place the environment first and the system second). Then, in a tensor product basis in which $|e\rangle \equiv |e_1\rangle$, the $(d \times d)$

matrices E_k form the first block column:

$$U = \left[\begin{array}{c|c|c} E_1 & \cdots & \cdot \\ \hline E_2 & \cdots & \cdot \\ \hline \vdots & \ddots & \vdots \\ \hline E_n & \cdots & \cdot \end{array}\right].$$

The remaining entries of U are irrelevant for the purpose of describing our sub-dynamics, since we have fixed the initial state of the environment. If we had chosen $|e\rangle \equiv |e_2\rangle$ instead, then we would have been looking at the matrices in the *second* block column, and all other entries would be irrelevant. You see the pattern, right? Here, the k in E_k ranges from 1 to some $n \leqslant D$, if we discount zero matrices. Finally, to be clear, reversing the order of the system and the environment does not affect the physical description *as long as we remember which part is which*. We changed the order just for pedagogical purposes; you can stick to the original order if you wish, but then the entries of E_k will be dispersed all over the matrix U.

Now, if we express the system-plus-environment evolution in terms of density operators

$$|\psi\rangle\langle\psi| \otimes |e\rangle\langle e| \longmapsto \sum_{k,l} E_k|\psi\rangle\langle\psi|E_l^\dagger \otimes |e_k\rangle\langle e_l|$$

and trace over the environment, then we can write the evolution of the system alone as

$$|\psi\rangle\langle\psi| \longmapsto \sum_k E_k|\psi\rangle\langle\psi|E_k^\dagger.$$

Recall that any density operator can be represented as a statistical mixture of pure states, and so:

The above equation can be generalised and written as

$$\rho \longrightarrow \sum_k E_k \rho E_k^\dagger$$

where ρ is any state of the system, pure or mixed.
It turns out that *all* quantum channels can be represented in this way, which is known as the **Kraus representation**, or the **operator sum representation**. In this context, the operators E_k are referred to as the **Kraus operators**.

Karl Kraus (1938–1988) was a German physicist.

The Kraus representation follows from the unitary representation once we choose an orthonormal basis in $\mathcal{H}_E$. Conversely, given a Kraus representation $E_k \in \mathscr{B}(\mathcal{H})$ of the channel $\mathcal{E}$, we can always view it as an operation resulting from a unitary map acting on some extended system: we just introduce an auxiliary system, the environment, which has dimension equal to the number of Kraus operators, and then choose an orthonormal basis $|e_k\rangle$. We then define U via $U|\psi\rangle|e\rangle \mapsto \sum_k E_k|\psi\rangle|e_k\rangle$. It is convenient to choose $|e\rangle \equiv |e_1\rangle$; in this case the operators E_k form the first block column of U. All other entries can be chosen as you please (as long as U remains unitary).

By now it should be clear that a Kraus representation of a given quantum channel is *not* unique: once we fix U and $|e\rangle$, and hence define a quantum channel, we may still choose the basis vectors $|e_k\rangle$ in $\mathcal{H}_E$ as we please. This means that the Kraus operators

$E_k = \langle e_k|U|e\rangle$ and $F_k = \langle f_k|U|e\rangle$, where $\{|e_k\rangle\}$ and $\{|f_k\rangle\}$ are two orthonormal bases, describe the same quantum channel. They are unitarily related:

$$\begin{aligned} F_k &= \langle f_k|U|e\rangle \\ &= \sum_l \langle f_k|e_l\rangle\langle e_l|U|e\rangle \\ &= \sum_l \langle f_k|e_l\rangle E_l \\ &= \sum_l R_{kl} E_l \end{aligned}$$

where $R_{kl} = \langle f_k|e_l\rangle$ is a unitary matrix, and we have used the decomposition of the identity $\sum_l |e_l\rangle\langle e_l| = \mathbf{1}_E$. In fact, the converse holds true as well:

> Suppose that $E_1, \ldots, E_n$ and $F_1, \ldots, F_m$ are Kraus operators associated to the quantum channels $\mathcal{E}$ and $\mathcal{F}$, respectively. Let us append zero operators to the shorter list to ensure that $n = m$.
> *Then $\mathcal{E}$ and $\mathcal{F}$ describe the same channel if and only if $F_k = \sum_l R_{kl} E_l$ for some unitary R.*

It is easy to see the reverse implication:

$$\begin{aligned} \mathcal{F}(\rho) &= \sum_k F_k \rho F_k^\dagger \\ &= \sum_{ijk} R_{ki} E_i \rho E_j^\dagger R_{jk}^\star \\ &= \sum_{ij} \underbrace{\left(\sum_k R_{jk}^\star R_{ki}\right)}_{\delta_{ij}} E_i \rho E_j^\dagger \\ &= \sum_j E_j \rho E_j^\dagger \\ &= \mathcal{E}(\rho). \end{aligned}$$

Without any restrictions on the operators E_k, any map of the form $U \mapsto \sum_k E_k U E_k^\dagger$ preserves Hermiticity and positivity. If we additionally require the completeness relation $\sum_k E_k^\dagger E_k = \mathbf{1}$, then the map becomes trace-preserving:

Prove this!

$$\begin{aligned} \operatorname{tr}\mathcal{E}(\rho) &= \operatorname{tr}\sum_k E_k \rho E_k^\dagger \\ &= \operatorname{tr}\left[\underbrace{\left(\sum_k E_k^\dagger E_k\right)}_{\mathbf{1}} \rho\right] \\ &= \operatorname{tr}\rho \\ &= 1. \end{aligned}$$

Any unitary evolution of a density operator

$$\rho \longmapsto U\rho U^\dagger$$

is a trivial example of the Kraus representation, with just one Kraus operator U. Next we may consider a probability distribution on a finite set of unitary operations $U_1, \ldots, U_m$. In this scenario, U_k will be chosen at random according to the probability distribution $p_1 \ldots p_k \ldots p_m$, and applied to ρ. If this is all that is known about the procedure, then the resulting transformation of ρ can be written as

$$\begin{aligned}\rho &\longmapsto \sum_k p_k U_k \rho U_k^\dagger \\ &= \sum_k \underbrace{\sqrt{p_k}U_k}_{E_k} \rho \underbrace{U_k^\dagger \sqrt{p_k}}_{E_k^\dagger} \\ &= \sum_k E_k \rho E_k^\dagger.\end{aligned}$$

In this case,

$$\rho \longmapsto \sum_k E_k \rho E_k^\dagger$$

where $\sum_k E_k^\dagger E_k = \sum_k E_k E_k^\dagger = \mathbf{1}$. This is an example of a **unital channel**, i.e. one for which $\sum_k E_k E_k^\dagger = \mathbf{1}$; such a channel sends maximally mixed inputs to maximally mixed outputs (if $\rho \propto \mathbf{1}$, then so too is ρ').

In particular, if we restrict our attention to qubits, and choose the unitaries to be the identity and the three Pauli operators, then we construct what is known as a **Pauli channel**:

$$\rho \longmapsto p_0 \mathbf{1}\rho\mathbf{1} + p_x X\rho X + p_y Y\rho Y + p_z Z\rho Z.$$

The most popular Pauli channel is the **depolarising channel**

$$\rho_p \longmapsto (1-p)\mathbf{1}\rho\mathbf{1} + \frac{p}{3}(X\rho X + Y\rho Y + Z\rho Z)$$

for $0 \leqslant p \leqslant 1$. In the depolarising channel, a qubit in state ρ_p remains intact with probability $1-p$, or is otherwise transformed under one of the Pauli operators X, Y, or Z, each chosen randomly with probability $p/3$. We study this channel in more detail in the exercises.

10.1.3 Expand, evolve, and discard

We shall now discuss two important operations: *adding* and *discarding* subsystems. You need to pay attention to the dimensions of the Hilbert spaces involved, for they are different. Given a system $\mathcal{A}$, we can bring in another system $\mathcal{B}$ to form the combined system $\mathcal{AB}$. This will be described by a map (**expansion**)

$$\mathscr{B}(\mathcal{H}_\mathcal{A}) \longrightarrow \mathscr{B}(\mathcal{H}_\mathcal{A} \otimes \mathcal{H}_\mathcal{B}).$$

Conversely, given a composite system $\mathcal{AB}$, we can discard subsystem $\mathcal{B}$, which is described by a map (**reduction**)

$$\mathscr{B}(\mathcal{H}_\mathcal{A} \otimes \mathcal{H}_\mathcal{B}) \longrightarrow \mathscr{B}(\mathcal{H}_\mathcal{A}).$$

We will show now that the two operations can be written in the Kraus representation. For consistency with the rest of the exposition, we will choose $\mathcal{A}$ to be the system and $\mathcal{B}$ the environment.

- **Expansion.** Any quantum system can be expanded by bringing in an auxiliary system (here, the environment) in a fixed state. This is described by the **isometric embedding**

  $$|\psi\rangle \longmapsto |\psi\rangle \otimes |e\rangle = (\mathbf{1} \otimes |e\rangle)|\psi\rangle.$$

  It takes vectors in the Hilbert space associated with the original system, $\mathcal{H}$ (where $\dim \mathcal{H} = d$) and tensors them with a fixed vector $|e\rangle$ in the Hilbert space associated with the environment, $\mathcal{H}_E$ (where $\dim \mathcal{H}_E = d_E$). So it *embeds* the structure of $\mathcal{H}$ into a larger Hilbert space $\mathcal{H} \otimes \mathcal{H}_E$ (where $\dim \mathcal{H} \otimes \mathcal{H}_\mathcal{A} = d \cdot d_E$). In terms of density operators, we write the expansion as

  $$\begin{aligned}\rho \longmapsto & \rho \otimes |e\rangle\langle e| \\ = & (\mathbf{1} \otimes |e\rangle)\rho(\mathbf{1} \otimes \langle e|) \\ = & E\rho E^\dagger\end{aligned}$$

  where $E = \mathbf{1} \otimes |e\rangle$ can be represented by a $(d \times (d \cdot d_E))$ matrix. We note that $E^\dagger E = \mathbf{1} \otimes \langle e|e\rangle = \mathbf{1}$ is the identity in $\mathcal{H}$.

- **Reduction.** We can also discard a part of a composed system. If $\rho \in \mathscr{B}(\mathcal{H} \otimes \mathcal{H}_E)$ is a joint state of the system and the environment (the two may be entangled), then we can neglect or discard the environment by tracing over it:

  $$\begin{aligned}& \rho \longmapsto \operatorname{tr}_E \rho \\ = & \sum_k (\mathbf{1} \otimes \langle e_k|)\rho(\mathbf{1} \otimes |e_k\rangle) \\ = & \sum_k E_k \rho E_k^\dagger\end{aligned}$$

  where $E_k = \mathbf{1} \otimes \langle e_k|$ and can be represented by a $(d \cdot d_E) \times d$ matrices. Again, we note that $\sum_k E_k^\dagger E_k = \mathbf{1} \otimes \sum_k |e_k\rangle\langle e_k| = \mathbf{1} \otimes \mathbf{1}_E$ is the identity in $\mathcal{H} \otimes \mathcal{H}_E$.

So both expansion and reduction can be expressed in the Kraus representation as $\rho \mapsto \sum_k E_k \rho E_k^\dagger$, where $\sum_k E_k^\dagger E_k = \mathbf{1}$. Here, the identity operator acts on the Hilbert space associated with the original density operator ρ, that is, $\mathcal{H}$ for the expansion and $\mathcal{H} \otimes \mathcal{H}_E$ for the reduction. Note that we do *not* require that $\sum_k E_k E_k^\dagger = \mathbf{1}$.

The next thing to notice is that we can compose quantum channels. This is straightforward in the Kraus representation: if

$$\begin{aligned}& \mathcal{E}_A \colon \rho \longmapsto \sum_i A_i \rho A_i^\dagger \\ & \text{where } \sum_i A_i^\dagger A_i = \mathbf{1}\end{aligned}$$

and

$$\mathcal{E}_B \colon \rho \longmapsto \sum_j B_j \rho B_j^\dagger$$
$$\text{where } \sum_j B_j^\dagger B_j = \mathbf{1}$$

then

$$\mathcal{E}_B \mathcal{E}_A \colon \rho \longmapsto \sum_{ij} (B_j A_i) \rho (B_j A_i)^\dagger$$

where the $B_j A_i$ are the Kraus operators associated with the new channel $\mathcal{E}_B \mathcal{E}_A$. We can verify that

$$\begin{aligned} \sum_{ij} (B_j A_i)^\dagger (B_j A_i) &= \sum_i A_i^\dagger \left(\sum_j B_j^\dagger B_j \right) A_i \\ &= \sum_i A_i^\dagger A_i \\ &= \mathbf{1}. \end{aligned}$$

We now postulate that *any quantum channel is a composition of the three elementary operations*:

1. **expansion**, i.e. bringing in another physical system in a fixed state (via an isometric embedding);
2. **unitary evolution** of the *whole* system; and
3. **reduction**, i.e. discarding everything we do not care about (via the partial trace).

Given that any unitary evolution $\rho \mapsto U \rho U^\dagger$ can be explicitly expressed in the operator sum form, the typical sequence "expand – unitary – reduce", which is the composition of three quantum channels, also admits an Kraus representation.

From a *mathematical* perspective, we have extended the class of quantum evolutions from

$$\rho \longmapsto U \rho U^\dagger \quad \text{with } U^\dagger U = \mathbf{1}$$

to

$$\rho \longmapsto \sum_k E_k \rho E_k^\dagger \quad \text{with } \sum_k E_k^\dagger E_k = \mathbf{1}.$$

To be clear, we are not claiming here that the Kraus representation has any fundamental significance: it simply follows from a unitary evolution of a larger quantum system, and it is a convenient mathematical shortcut when we deal with subsystems of that larger system.

10.2 The axiomatic approach

10.2.1 Completely positive maps

Let us now turn to the axiomatic approach. What are the most general transformations that map density operators to density operators?

The first thing to notice is that any such a transformation $\mathcal{E}$ must respect the mixing of states. Consider an ensemble of systems, with a fraction p_1 of them in the state ρ_1, and the remaining p_2 of them in the state ρ_2. The overall ensemble is described by $\rho = p_1\rho_1 + p_2\rho_2$. If we apply $\mathcal{E}$ to each member of the ensemble, then the overall ensemble will be described by the density operator $\rho' = \mathcal{E}(\rho)$, which is $\rho' = p_1\mathcal{E}(\rho_1) + p_2\mathcal{E}(\rho_2)$. We conclude that $\mathcal{E}$ must be a *linear* map on density operators: $\mathcal{E}\colon \mathscr{B}(\mathcal{H}) \mapsto \mathscr{B}(\mathcal{H}')$. Moreover, since $\mathcal{E}$ maps density operators to density operators, it must be **positive** ($\mathcal{E}(\rho) \geqslant 0$ whenever $\rho \geqslant 0$) and **trace preserving** ($\operatorname{tr}\mathcal{E}(\rho) = \operatorname{tr}\rho$ for all $\rho \in \mathscr{B}(\mathcal{H})$). Can we stop here and conclude that quantum channels are described by linear, positive, trace-preserving maps?

More precisely, we require that $\mathcal{E}$ be an *affine* map.

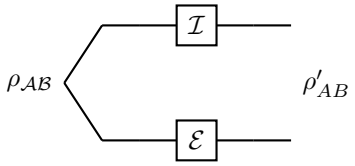
It turns out that there is one more property which must added to the list: we must require that the map remains positive even when it acts on a part of a larger system. That is, for any bipartite state $\rho_{\mathcal{AB}}$ of a composed system $\mathcal{AB}$, the extended map $\mathcal{I}\otimes\mathcal{E}$ (meaning "apply $\mathcal{E}$ to either one of the subsystems (it does not matter which one; here we have chosen subsystem $\mathcal{B}$), and do nothing (i.e. apply the identity map $\mathcal{I}$) to the remaining subsystem") gives a proper density operator ρ'_{AB}:

$$\rho'_{AB} = (\mathcal{I}\otimes\mathcal{E})(\rho_{\mathcal{AB}}) \geqslant 0.$$

Note the difference between $\mathbf{1}$ and $\mathcal{I}$. They are both identity maps, but $\mathbf{1}$ is the identity operator on $\mathcal{H}$, i.e. the (unitary) identity matrix; $\mathcal{I}$ is the identity *superoperator* (or quantum channel), i.e. the identity superoperator that acts on operators on $\mathcal{H}$.

In fact, a *lot* (if not *all*) of quantum information theory can be converted into similar looking diagrams, using the formalism of **string diagrams in monoidal categories**. Another graphical formalism is that of **ZX-calculus**. Both of these topics can lead into very deep rabbit holes very quickly, so beware!

We can represent this construction of ρ'_{AB} diagrammatically as follows:



Now this is a stronger property than mere positivity; now we are asking for something called **complete positivity**. You may naively think that such trivial extensions of positive maps must be also positive, but this is not the case: the positivity of $\mathcal{E}$ does *not* imply the positivity of its extension $\mathcal{I}\otimes\mathcal{E}$.

> *Positivity does not imply complete positivity!* Why? *Because physical systems do not exist in isolation*; applying $\mathcal{E}$ to a part of the whole and doing nothing (i.e. applying the identity) to the rest may mess up the state of the whole in an unacceptable way.

This is a very quantum phenomenon, and is another consequence of quantum entanglement between parts of the composite system.

A simple example of a positive map that is not completely positive is the transpose. Consider the transpose operation on a single qubit: $T\colon |i\rangle\langle j| \mapsto |j\rangle\langle i|$ (for $i,j = 0,1$). It preserves both trace and positivity, and, if ρ is a density matrix, then so is $\rho' = \rho^T$. However, if the input qubit is part of a *two-qubit* system, initially in the entangled state $|\Omega\rangle = \frac{1}{\sqrt{2}}(|0\rangle|0\rangle + |1\rangle|1\rangle)$, and the transpose is applied to *only one* of the two qubits, (say, the second one) then, under the action of the partial transpose $\mathcal{I}\otimes T$, the density matrix

of the two qubits becomes

$$\begin{aligned}|\Omega\rangle\langle\Omega| =& \frac{1}{2}\sum_{ij}|i\rangle\langle j|\otimes|i\rangle\langle j| \\ \overset{\mathcal{I}\otimes T}{\longmapsto}& \frac{1}{2}\sum_{ij}|i\rangle\langle j|\otimes T(|i\rangle\langle j|) \\ =& \frac{1}{2}\sum_{ij}|i\rangle\langle j|\otimes|j\rangle\langle i|.\end{aligned}$$

The output is the SWAP matrix, since it describes the SWAP operation: $|j\rangle|i\rangle \mapsto |i\rangle|j\rangle$. Since the square of SWAP is the identity, its eigenvalues are ± 1; states which are symmetric under interchange of the two qubits have eigenvalue 1, while antisymmetric states have eigenvalue -1. Thus the SWAP matrix has negative eigenvalues, which means that $\mathcal{I}\otimes T$ does *not* preserve positivity, and therefore T is *not* a completely positive map. If you prefer to see this more explicitly, then you can use the matrix representation of $|\Omega\rangle\langle\Omega|$, apply the partial transpose $\mathcal{I}\otimes T$, and inspect the resulting matrix:

$$\frac{1}{2}\left[\begin{array}{cc|cc} 1&0&0&1\\0&0&0&0\\\hline 0&0&0&0\\1&0&0&1\end{array}\right] \overset{\mathcal{I}\otimes T}{\longmapsto} \frac{1}{2}\left[\begin{array}{cc|cc} 1&0&0&0\\0&0&1&0\\\hline 0&1&0&0\\0&0&0&1\end{array}\right].$$

The partial transpose maps the density matrix of a maximally entangled state $|\Omega\rangle\langle\Omega|$ to the SWAP matrix, which has one negative eigenvalue (namely -1), and so is not a density matrix.

So how can we tell if a given map $\mathcal{E}$ is completely positive or not? The answer is "in pretty much the same way we discovered that the transpose is not a completely positive map". It suffices to apply $\mathcal{I}\otimes\mathcal{E}$ (or $\mathcal{E}\otimes\mathcal{I}$) to the maximally entangled state of two subsystems and check if the result is a legal density matrix. In fact, it works both ways:

> A linear map $\mathcal{E}\colon B(\mathcal{H}) \to B(\mathcal{H}')$ is completely positive if and only if $(\mathcal{I}\otimes\mathcal{E})(|\Omega\rangle\langle\Omega|) \geqslant 0$, where $|\Omega\rangle$ is a maximally entangled state of dimension $d = \dim\mathcal{H}$.

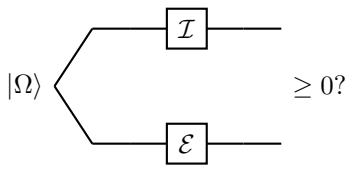


Figure 39. Test for complete positivity: apply the map $\mathcal{E}$ to half of a maximally entangled state and check if the resulting bipartite operator is a density operator.

Clearly, if $\mathcal{E}$ is completely positive, then, from the definition, $(\mathcal{I}\otimes\mathcal{E})(|\Omega\rangle\langle\Omega|) \geqslant 0$. Conversely, in order to check if $(\mathcal{I}\otimes\mathcal{E})(\rho_{\mathcal{AB}}) \geqslant 0$ holds for any bipartite state $\rho_{\mathcal{AB}}$, it is sufficient to consider only pure states, e.g. the states that appear in the spectral decomposition of $\rho_{\mathcal{AB}}$. Any such pure bipartite state $|\Psi\rangle$ can be written as $(A\otimes\mathbf{1})|\Omega\rangle$ or $(\mathbf{1}\otimes B)|\Omega\rangle$, for some linear operators A and B, on subsystems $\mathcal{A}$ or $\mathcal{B}$, respectively:

The state $|\Psi\rangle$ has the Schmidt decomposition $\sum_k \lambda_k|k\rangle|k\rangle$, where λ_k^2 are the eigenvalues of the reduced density operators σ_A and σ_B. Thus $|\Psi\rangle$ can be written as $(\sqrt{d\sigma_A}\otimes\mathbf{1})|\Omega\rangle$ or $(\mathbf{1}\otimes\sqrt{d\sigma_B})|\Omega\rangle$, where d is the Schmidt rank.

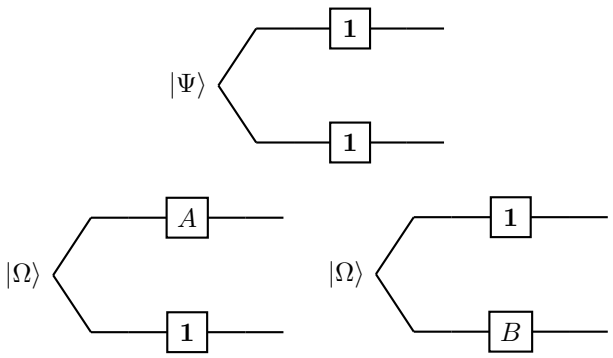


Figure 40. All three of these operations are equal.

Given that the map $\mathcal{E}$ acts only on the subsystem $\mathcal{B}$, we have that

$$\begin{aligned}&(\mathcal{I}\otimes\mathcal{E})[(A\otimes\mathbf{1})|\Omega\rangle\langle\Omega|(A^\dagger\otimes\mathbf{1})]\\&=(A\otimes\mathbf{1})[(\mathcal{I}\otimes\mathcal{E})(|\Omega\rangle\langle\Omega|)](A^\dagger\otimes\mathbf{1}),\end{aligned}$$

and if $(\mathcal{I}\otimes\mathcal{E})(|\Omega\rangle\langle\Omega|)$ is positive semidefinite, then so too is $(A\otimes\mathbf{1})[(\mathcal{I}\otimes\mathcal{E})(|\Omega\rangle\langle\Omega|)](A^\dagger\otimes\mathbf{1})$.

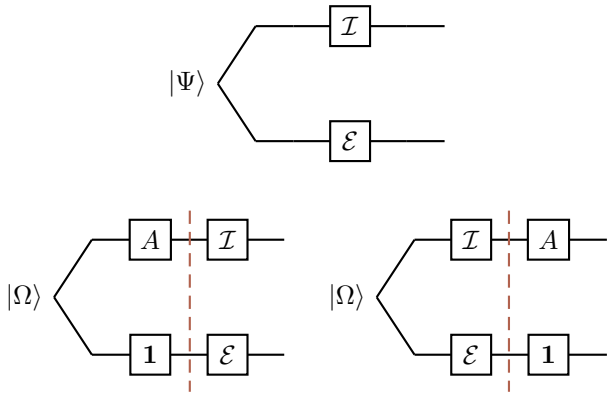


Figure 41. All three of these operations are also equal.

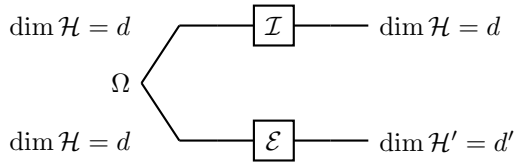
10.2.2 The Choi matrix

If we choose bases of $\mathcal{H}$ and $\mathcal{H}'$ (where $\dim\mathcal{H}=d$ and $\dim\mathcal{H}'=d'$), then any linear map $\mathcal{E}\colon\mathscr{B}(\mathcal{H})\to\mathscr{B}(\mathcal{H}')$ can be completely characterised by its action on the d^2 basis matrices $|i\rangle\langle j|$, where $i,j=1,2\ldots,d$.

The **Choi matrix** $\widetilde{\mathcal{E}}$ of $\mathcal{E}$ is defined by

$$\widetilde{\mathcal{E}} = (\mathcal{I} \otimes \mathcal{E})|\Omega\rangle\langle\Omega| = \frac{1}{d}\sum_{i,j} |i\rangle\langle j| \otimes \mathcal{E}(|i\rangle\langle j|)$$

where $|\Omega\rangle = \sum_i |i\rangle|i\rangle$ is a maximally entangled state.



The Choi matrix is a neat way of tabulating all possible values of $\mathcal{E}(|i\rangle\langle j|)$, since the (i,j)-th block entry of $\widetilde{\mathcal{E}}$ is the $(d' \times d')$-matrix $\mathcal{E}(|i\rangle\langle j|)$:

$$\widetilde{\mathcal{E}} = \begin{bmatrix} \mathcal{E}(|0\rangle\langle 0|) & \mathcal{E}(|0\rangle\langle 1|) & \mathcal{E}(|0\rangle\langle 2|) & \cdots \\ \mathcal{E}(|1\rangle\langle 0|) & \mathcal{E}(|1\rangle\langle 1|) & \mathcal{E}(|1\rangle\langle 0|) & \cdots \\ \mathcal{E}(|2\rangle\langle 0|) & \mathcal{E}(|2\rangle\langle 1|) & \mathcal{E}(|2\rangle\langle 2|) & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}.$$

Of course, once we know how $\mathcal{E}$ acts on the d^2 basis matrices $|i\rangle\langle j|$, we know how it acts on any $(d \times d)$ matrix $X = \sum_{ij} X_{ij}|i\rangle\langle j|$, since $\mathcal{E}(X) = \sum_{ij} X_{ij}\mathcal{E}(|i\rangle\langle j|)$, where $i, j = 1, \ldots, d$. This can be written as follows:

$$\mathcal{E}(X) = d\,\mathrm{tr}_{\mathcal{A}}[(X^T \otimes \mathbf{1})\widetilde{\mathcal{E}}]$$

where the partial trace is taken over the "first half" of the entangled state, the part to which we did not apply $\mathcal{E}$ in our construction of $\widetilde{\mathcal{E}}$.

In practice it is more convenient to use the following formula (which holds for any $(d \times d)$ matrix X and any $(d' \times d')$ matrix Y):

$$\mathrm{tr}[\mathcal{E}(X)Y] = \mathrm{tr}[\widetilde{\mathcal{E}}(X^T \otimes Y)]$$

For example, if we are interested in the component $\mathcal{E}(X)_{ij} = \langle i|\mathcal{E}(X)|j\rangle$, then we take $Y = |j\rangle\langle i|$.

The Choi matrix $\widetilde{\mathcal{E}}$ in $\mathscr{B}(\mathcal{H}\otimes\mathcal{H}')$ is essentially another way of representing the linear map $\mathcal{E}\colon \mathscr{B}(\mathcal{H}) \to \mathscr{B}(\mathcal{H}')$. One can express all properties of $\mathcal{E}$ in terms of properties of $\widetilde{\mathcal{E}}$, e.g.

- $\mathcal{E}$ is completely positive if and only if $\widetilde{\mathcal{E}} \geqslant 0$;
- $\mathcal{E}$ is trace preserving if and only if $\mathrm{tr}_{\mathcal{B}}\,\widetilde{\mathcal{E}} = d\mathbf{1}$".

We now claim that *all quantum channels are described by completely positive trace preserving (CPTP) maps*. This is simply because we cannot think of any further restrictions we should be imposing on the linear maps $\mathscr{B}(\mathcal{H}) \to \mathscr{B}(\mathcal{H}')$ to make them physically admissible. Needless to say, mathematics alone cannot tell us what is and what is not physically admissible, but we can certainly compare this axiomatic approach with the constructive one. As it happens, our claims makes sense, since we can show that a map is completely positive if and only if it can be written in the operator sum form. In other words, *quantum channels are exactly completely positive trace-preserving (CPTP) maps*. We will prove this in the next section

As a closing remark before moving on, note that the equivalence of linear maps $\mathscr{B}(\mathcal{H}) \to \mathscr{B}(\mathcal{H}')$ and matrices in $\mathscr{B}(\mathcal{H} \otimes \mathcal{H}')$ is well known. In the context of quantum channels it is often referred to as the **Choi–Jamiokowski isomorphism**, or as **channel–state duality** (which we explain in more detail later on). For CPTP maps, the Choi matrix is a density matrix, also known as the **Jamiokowski state**, and the correspondence is referred to as the **channel-state duality**. Mathematically, it is hardly surprising that the matrix elements of an operator on a tensor product can be reorganised and reinterpreted as the matrix elements of an operator between operator spaces. What *is* interesting, and perhaps not so obvious, is that the positivity conditions for maps and their Choi matrices match up exactly under this correspondence.

10.2.3 An example

Any linear map $\mathcal{E}$ acting on a qubit can be completely characterised by its action on the four basis matrices $|i\rangle\langle j|$ (for $i, j = 0, 1$) and thus represented as the (4×4) Choi matrix

$$\widetilde{\mathcal{E}} = \frac{1}{2}\left[\begin{array}{c|c} \mathcal{E}(|0\rangle\langle 0|) & \mathcal{E}(|0\rangle\langle 1|) \\ \hline \mathcal{E}(|1\rangle\langle 0|)) & \mathcal{E}(|1\rangle\langle 1|) \end{array}\right].$$

As an example, consider the following map:

$$\mathcal{E}_p\colon |i\rangle\langle j| \longmapsto p|j\rangle\langle i| + \delta_{ij}\frac{(1-p)}{2}|i\rangle\langle j|,$$

where $0 \leqslant p \leqslant 1$ is a real parameter. Its action on the four basis matrices is

$$
\begin{aligned}
|0\rangle\langle 0| = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} &\longmapsto \begin{bmatrix} \frac{1+p}{2} & 0 \\ 0 & \frac{1-p}{2} \end{bmatrix} \\
|0\rangle\langle 1| = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} &\longmapsto \begin{bmatrix} 0 & 0 \\ p & 0 \end{bmatrix}, \\
|1\rangle\langle 0| = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} &\longmapsto \begin{bmatrix} 0 & p \\ 0 & 0 \end{bmatrix} \\
|1\rangle\langle 1| = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} &\longmapsto \begin{bmatrix} \frac{1-p}{2} & 0 \\ 0 & \frac{1+p}{2} \end{bmatrix}
\end{aligned}
$$

which we can neatly collect and tabulate in the corresponding Choi matrix:

$$
\widetilde{\mathcal{E}}_p = \frac{1}{2} \left[\begin{array}{cc|cc} \frac{1+p}{2} & 0 & 0 & 0 \\ 0 & \frac{1-p}{2} & p & 0 \\ \hline 0 & p & \frac{1-p}{2} & 0 \\ 0 & 0 & 0 & \frac{1+p}{2} \end{array}\right].
$$

This matrix contains *all* the information concerning the map $\mathcal{E}_p$. For example, $\mathcal{E}_p$ is completely positive if and only if $\widetilde{\mathcal{E}}_p \geqslant 0$, which happens for $p \leqslant \frac{1}{3}$ (the eigenvalues of $\widetilde{\mathcal{E}}_p$ are $\frac{1}{4}(1+p)$ and $\frac{1}{4}(1-3p)$).

Now the map $\mathcal{E}_p$ sends a density operator ρ to

$$
\rho \longmapsto p\rho^T + \frac{(1-p)}{2}\mathbf{1},
$$

where ρ^T is the transpose of ρ. Is this map a quantum channel? That is, does it represent a physical process that can be implemented in a lab? In the expression above, the convex sum on the right hand side can be interpreted as follows: take the input state ρ and either apply the transpose or replace it with the maximally mixed state, with probabilities p and $1-p$ respectively. This is fine, except that the transpose operation is *not* completely positive, and as such it is not physically admissible; it cannot be implemented. Does it mean that this map cannot be implemented? No, it does not: it just depends on the value of p. The case $p = 0$ corresponds to just replacing the input with the maximally mixed state, which is something that can be easily implemented. However, as p increases from 0 to 1, at some critical point the map switches from completely positive to positive. We know from the previous example that this critical point corresponds to $p = \frac{1}{3}$.

10.3 Comparing the two approaches

The one remaining thing (well, almost) left to do is to bring the Kraus representation approach and the completely positive trace-preserving maps approach together, and show that they are essentially the same thing.

A linear map $\mathcal{E}\colon \mathscr{B}(\mathcal{H}) \to \mathscr{B}(\mathcal{H}')$ is completely positive if and only if it admits a Kraus decomposition of the form

$$\mathcal{E}(\rho) = \sum_k E_k \rho E_k^\dagger.$$

When this is the case, the above decomposition has the following properties:

- $\mathcal{E}$ is trace preserving if and only if $\sum_k E_k^\dagger E_k = \mathbf{1}$.
- Two sets of Kraus operators $\{E_k\}$ and $\{F_l\}$ represent the same map $\mathcal{E}$ if and only if there is a unitary R such that $E_k = \sum_l R_{kl} F_l$, where the smaller set of the Kraus operators is padded with zeros.
- For *any* $\mathcal{E}\colon \mathscr{B}(\mathcal{H}) \to \mathscr{B}(\mathcal{H}')$, there is *always* a representation with at most dd' mutually orthogonal Kraus operators: $\operatorname{tr} E_i^\dagger E_j \propto \delta_{ij}$.

Any map $\mathcal{E}$ written in the Kraus form is completely positive, since its Choi matrix $\widetilde{\mathcal{E}}$ is positive semidefinite:

$$\begin{aligned}\widetilde{\mathcal{E}} &= \mathbf{1} \otimes \mathcal{E}|\Omega\rangle\langle\Omega| \\ &= \sum_k (\mathbf{1} \otimes E_k)|\Omega\rangle\langle\Omega|(\mathbf{1} \otimes E_k^\dagger) \\ &\geqslant 0.\end{aligned}$$

Here, the Choi matrix is of the form $\sum_k |v_k\rangle\langle v_k|$, where $|v_k\rangle = (\mathbf{1} \otimes E_k)|\Omega\rangle$. This is a positive semidefinite matrix since, for any vector $|\Psi\rangle$, the expression $\sum_k \langle\Psi|v_k\rangle\langle v_k|\Psi\rangle$ is non-negative.

Conversely, if $\mathcal{E}$ is completely positive, then its Choi matrix is positive semidefinite, which means that it can be written as a mixture of pure states $|\psi_k\rangle\langle\psi_k|$ with probabilities p_k:

$$\widetilde{\mathcal{E}} = \sum_k |\widetilde{\psi_k}\rangle\langle\widetilde{\psi_k}|$$

where $|\widetilde{\psi_k}\rangle = \sqrt{p_k}|\psi_k\rangle$ are (non-normalised) vectors, and any such vector can be written as

$$|\widetilde{\psi_k}\rangle = (\mathbf{1} \otimes E_k)|\Omega\rangle \tag{10.3.1}$$

for some operator E_k. Hence

$$\begin{aligned}\widetilde{\mathcal{E}} &= \sum_k |\widetilde{\psi_k}\rangle\langle\widetilde{\psi_k}| \\ &= \sum_k (\mathbf{1} \otimes E_k)|\Omega\rangle\langle\Omega|(\mathbf{1} \otimes E_k^\dagger) \\ &= \frac{1}{d}\sum_{ij} |i\rangle\langle j| \otimes \underbrace{\sum_k E_k(|i\rangle\langle j|)E_k^\dagger}_{\mathcal{E}(|i\rangle\langle j|)}.\end{aligned}$$

Comparing the last expression on the right with the definition of $\widetilde{\mathcal{E}}$, we conclude that

$$\mathcal{E}(\rho) = \sum_k E_k \rho E_k^\dagger.$$

The unitary freedom of choosing the Kraus operators is directly related to the unitary freedom of choosing the statistical ensemble for $\widetilde{\mathcal{E}}$. We know that two mixtures $\{p_k, |\psi_k\rangle\}$ and $\{q_l, |\phi_l\rangle\}$ described by the same density operator

$$\widetilde{\mathcal{E}} = \sum_k |\widetilde{\psi_k}\rangle\langle\widetilde{\psi_k}| = \sum_l |\widetilde{\phi_l}\rangle\langle\widetilde{\phi_l}|,$$

where $|\widetilde{\psi_k}\rangle = \sqrt{p_k}|\psi_k\rangle$ and $|\widetilde{\phi_l}\rangle = \sqrt{q_l}|\phi_l\rangle$ are related via

$$|\widetilde{\psi_k}\rangle = \sum_l R_{kl}|\widetilde{\phi_l}\rangle,$$

for some unitary R. Given the correspondence in Equation (10.3.1), this implies the same unitary freedom in choosing the Kraus operators. The number of vectors contributing to each mixture, and hence the number of corresponding Kraus operators, may be different, so we simply extend the smaller set to the required size by adding zero operators.

We can also see that the minimal number of Kraus operators needed to express $\mathcal{E}$ in the Kraus form is given by the rank of its Choi matrix $\widetilde{\mathcal{E}}$, and so we need no more than dd' such operators. This minimal set of Kraus operators corresponds to the spectral decomposition of $\widetilde{\mathcal{E}}$. Indeed, if $\widetilde{\mathcal{E}} = \sum_k |\widetilde{v_k}\rangle\langle\widetilde{v_k}|$ and $|\widetilde{v_k}\rangle = (\mathbf{1}\otimes E_k)|\Omega\rangle$, then the orthogonality of $|\widetilde{v_k}\rangle$ and $|\widetilde{v_l}\rangle$ implies the orthogonality, in the Hilbert-Schmidt sense, of the corresponding Kraus operators E_k and E_l. In order to see this, we express $\langle\widetilde{v_k}|\widetilde{v_l}\rangle$ as

Indeed, any "square root" of the Choi matrix (i.e. a matrix B such that $B^\dagger B = \widetilde{\mathcal{E}}$) gives a set of Kraus operators by reading off the column vectors of B.

$$\begin{aligned}\langle\widetilde{v_k}|\widetilde{v_l}\rangle &= \langle\Omega|(\mathbf{1}\otimes E_k^\dagger)(\mathbf{1}\otimes E_l)|\Omega\rangle \\ &= \operatorname{tr}\left[(\mathbf{1}\otimes E_k^\dagger E_l)|\Omega\rangle\langle\Omega|\right] \\ &= \frac{1}{d}\operatorname{tr}\sum_{ij}|i\rangle\langle j|\otimes E_k^\dagger E_l|i\rangle\langle j|\end{aligned}$$

where we have substituted $\frac{1}{d}\sum_{ij}|i\rangle\langle j|\otimes|i\rangle\langle j|$ for $|\Omega\rangle\langle\Omega|$. The trace of the tensor product of two matrices is the product of their traces, whence

$$\begin{aligned}\langle\widetilde{v_k}|\widetilde{v_l}\rangle &= \frac{1}{d}\sum_{ij}\langle i|j\rangle\operatorname{tr}E_k^\dagger E_l|i\rangle\langle j| \\ &= \frac{1}{d}\operatorname{tr}E_k^\dagger E_l\end{aligned}$$

for $\langle i|j\rangle = \delta_{ij}$ and $\sum_i |i\rangle\langle i| = \mathbf{1}$. That is, $\langle\widetilde{v_k}|\widetilde{v_l}\rangle = 0$ implies that $\operatorname{tr}E_k^\dagger E_l = 0$.

10.4 What are positive maps good for?

Positive maps that are not completely positive are not completely useless. True, they cannot describe any quantum dynamics, but still they have useful applications — for example, they can help us to determine if a given state is entangled or not.

Recall: a quantum state of a bipartite system $\mathcal{AB}$ described by the density matrix $\varrho^{\mathcal{AB}}$ is said to be **separable** if $\varrho^{\mathcal{AB}}$ is of the form

$$\varrho^{\mathcal{AB}} = \sum_k p_k \rho_k^{\mathcal{A}}\otimes\rho_k^{\mathcal{B}}$$

where $p_k \geqslant 0$ and $\sum_{k=1} p_k = 1$; otherwise $\varrho^{\mathcal{AB}}$ is said to be **entangled**. If we apply the partial transpose $\mathbf{1}\otimes T$ to this state, then it remains separable, since, as we have seen, the transpose ρ^B is a legal density matrix.

Positive maps, such as the transpose, can be quite deceptive: you have to include other systems in order to detect their unphysical character.

In separable states, one subsystem does not really know about the existence of the other, and so applying a positive map to one part produces a proper density operator, and thus does *not* reveal the unphysical character of the map. So, for *any separable state* ρ, we have $(\mathbf{1} \otimes T)\rho \geqslant 0$.

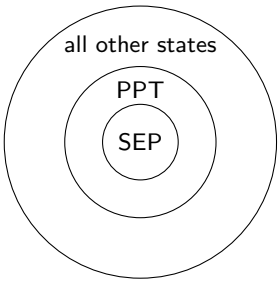
As an example, consider a quantum state of two qubits which is a mixture of the maximally entangled state $|\psi\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$ and the maximally mixed state described by the density matrix

$$\rho_p = p|\psi\rangle\langle\psi| + \frac{(1-p)}{4}\mathbf{1} \otimes \mathbf{1},$$

where $p \in [0, 1]$. If we apply the partial transpose $\mathbf{1} \otimes T$ to this state, and check for which values of p the resulting matrix is a density matrix, we see that, for all $p \in [\frac{1}{3}, 1]$, the density operator ρ describes an entangled state.

Note that the implication "if separable then the partial transpose is positive" does not imply the converse: there exist entangled states for which the partial transpose is positive, and they are known as the **entangled PPT states** However, for *two* qubits, the PPT states are exactly the separable states.

"PPT" stands for *positive partial transpose*.



10.5 Remarks and exercises

10.5.1 The depolarising channel

Show that the action of the depolarising channel

$$\rho_p \longmapsto (1-p)\mathbf{1}\rho\mathbf{1} + \frac{p}{3}(X\rho X + Y\rho Y + Z\rho Z)$$

is equivalent to the following operation: *with probability* p*, the input state is thrown away and replaced with the maximally mixed state, and with probability* $(1-p)$ *the input state survives without any disturbance*.

Hint. Use the Bloch sphere approach, and show that the depolarising channel uniformly contracts the sphere by the factor $1 - \frac{4}{3}p$.

10.5.2 The Choi–Jamiokowski isomorphism

The correspondence between linear maps $\mathscr{B}(\mathcal{H}) \to \mathscr{B}(\mathcal{H}')$ and operators in $\mathscr{B}(\mathcal{H} \otimes \mathcal{H}')$, known as the **Choi–Jamiokowski isomorphism** or **channel–state duality**, is another example of a well known correspondence between vectors in $\mathcal{H}_\mathcal{A} \otimes \mathcal{H}_\mathcal{B}$ and operators $\mathscr{B}(\mathcal{H}^\star_\mathcal{A}, \mathcal{H}_\mathcal{B})$ or $\mathscr{B}(\mathcal{H}^\star_\mathcal{B}, \mathcal{H}_\mathcal{A})$.

It is slightly confusing at first, but the **Choi isomorphism**, the **Jamiokowski isomorphism**, and the **Choi–Jamiokowski** isomorphism are really three distinct things: the first is very nice, but non-canonical (i.e. is dependent on the choice of basis); the second (for which I have no nice citation, but is basically given by considering $\sum |j\rangle\langle i| \otimes \mathcal{E}(|i\rangle\langle j|)$ instead of $\sum |i\rangle\langle j| \otimes \mathcal{E}(|i\rangle\langle j|)$) is canonical, but doesn't always map CP maps to

Take a tensor product vector in $|a\rangle \otimes |b\rangle \in \mathcal{H}_\mathcal{A} \otimes \mathcal{H}_\mathcal{B}$. Then it defines natural maps in $\mathscr{B}(\mathcal{H}_\mathcal{A}^\star, \mathcal{H}_\mathcal{B})$ and $\mathscr{B}(\mathcal{H}_\mathcal{B}^\star, \mathcal{H}_\mathcal{A})$, via

$$\begin{aligned}\langle x| &\longmapsto \langle x|a\rangle|b\rangle \\ \langle y| &\longmapsto |a\rangle\langle y|b\rangle\end{aligned}$$

for any linear forms $\langle x| \in \mathcal{H}_A^\star$ and $\langle y| \in \mathcal{H}_B^\star$. We then extend this construction (by linearity) to any vector in $\mathcal{H}_\mathcal{A} \otimes \mathcal{H}_\mathcal{B}$. These isomorphisms are **canonical**: they do not depend on the choice of any bases in the vectors spaces involved.

However, some care must be taken when we want to define correspondence between vectors in $\mathcal{H}_\mathcal{A} \otimes \mathcal{H}_\mathcal{B}$ and operators in $\mathscr{B}(\mathcal{H}_\mathcal{A}, \mathcal{H}_\mathcal{B})$ or $\mathscr{B}(\mathcal{H}_\mathcal{B}, \mathcal{H}_\mathcal{A})$. For example, physicists like to "construct" $\mathscr{B}(\mathcal{H}_\mathcal{B}, \mathcal{H}_\mathcal{A})$ in a deceptively simple way:

$$|a\rangle|b\rangle \longleftrightarrow |a\rangle\langle b|.$$

(where we have simply omitted the tensor product symbol). Flipping $|b\rangle$ and switching from $\mathcal{H}_\mathcal{B}$ to $\mathcal{H}_\mathcal{B}^\star$ is an *anti-linear* operation (since it involves complex conjugation). This is fine *when we stick to a specific basis $|i\rangle|j\rangle$ and use the ket-flipping approach only for the basis vectors*. This means that, for $|b\rangle = \sum_j \beta_j |j\rangle$, the correspondence looks like

$$|i\rangle|b\rangle \longleftrightarrow \sum_j \beta_j |i\rangle\langle j|$$

and *not* like

$$|i\rangle|b\rangle \longleftrightarrow |i\rangle\langle b| = \sum_j \beta_j^\star |i\rangle\langle j|.$$

This isomorphism is **non-canonical**: it depends on the choice of the basis. But it is still a pretty useful isomorphism! The Choi–Jamiokowski isomorphism is of this kind (i.e. non-canonical) — it works in the basis in which you express a maximally entangled state $|\Omega\rangle = \sum_i |i\rangle|i\rangle$.

10.5.3 Any bipartite state from a maximally entangled state

Any vector $|\psi\rangle$ in $\mathcal{H} \otimes \mathcal{H}$ can be written as

$$\begin{aligned}|\psi\rangle &= \sum_{ij} C_{ij} |i\rangle \otimes |j\rangle \\ &= \sum_j \underbrace{\left(\sum_i C_{ij}|i\rangle\right)}_{C|j\rangle} \otimes |j\rangle \\ &= \sum_i |i\rangle \otimes \underbrace{\left(\sum_j C_{ij}|j\rangle\right)}_{C^T|i\rangle}\end{aligned}$$

which we can write as

$$\begin{aligned}|\psi\rangle &= (C \otimes \mathbf{1})|\Omega\rangle \\ &= (\mathbf{1} \otimes C^T)|\Omega\rangle.\end{aligned}$$

The normalisation factor $\frac{1}{\sqrt{d}}$ can be incorporated into the matrix C, or you can just remember that it should be added at the end. Here

$$\begin{aligned} C|j\rangle &= \sum_i |i\rangle\langle i|C|j\rangle \\ &= \sum_i C_{ij}|i\rangle \end{aligned}$$

$$\begin{aligned} C^T|i\rangle &= \sum_j |j\rangle\langle j|C^T|i\rangle \\ &= \sum_j C_{ij}|j\rangle. \end{aligned}$$

10.5.4 Tricks with a maximally entangled state

A maximally entangled state of a bipartite system can be written, using the Schmidt decomposition, as

$$\begin{aligned} |\Omega\rangle &= \frac{1}{\sqrt{d}} \sum_i |i\rangle|i\rangle \\ |\Omega\rangle\langle\Omega| &= \frac{1}{d} \sum_{ij} |i\rangle\langle j| \otimes |i\rangle\langle j| \end{aligned}$$

Each subsystem is of dimension d, and all the Schmidt coefficients are equal. Here are few useful tricks involving a maximally entangled state.

- If we take the transpose in the Schmidt basis of $|\Omega\rangle$, then

  $$\langle\Omega|A \otimes B|\Omega\rangle = \frac{1}{d} \operatorname{tr}(A^T B).$$

- Any pure state of the bipartite system $|\psi\rangle = \sum_{ij} c_{ij}|i\rangle|j\rangle$ can be written as

  $$(C \otimes \mathbf{1})|\Omega\rangle = (\mathbf{1} \otimes C^T)|\Omega\rangle.$$

  This implies that

  $$(U \otimes \overline{U})|\Omega\rangle = |\Omega\rangle$$

  (where $\overline{U}$ denotes the matrix given by taking the complex conjugate, entry-wise, of U, i.e. *without* also taking the transpose).

- The swap operation, $S\colon |i\rangle|j\rangle \mapsto |j\rangle|i\rangle$, can be expressed as

  We write $X^{T_{\mathcal{A}}}$ to mean the partial transpose over $\mathcal{A}$, i.e. $T \otimes \mathcal{I}$.

  $$\begin{aligned} S &= d|\Omega\rangle\langle\Omega|^{T_{\mathcal{A}}} \\ &= d \sum_{ij} \big(|i\rangle\langle j|\big)^T \otimes |i\rangle\langle j| \\ &= d \sum_{ij} |j\rangle\langle i| \otimes |i\rangle\langle j|. \end{aligned}$$

  This implies that

  $$\operatorname{tr}[(A \otimes B)S] = \operatorname{tr} AB$$

and that

$$(A \otimes \mathbf{1})S = S(\mathbf{1} \otimes A).$$

10.5.5

Define the map $\mathcal{E}$ on a single qubit via

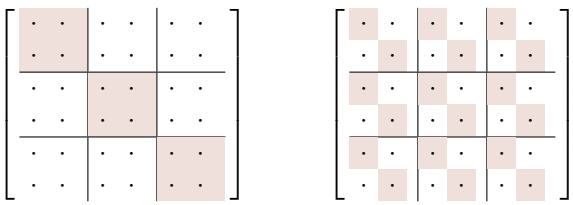
$$\begin{aligned}\mathcal{E}\colon \mathbf{1} &\longmapsto \mathbf{1}\\ \sigma_x &\longmapsto x\sigma_x\\ \sigma_y &\longmapsto y\sigma_y\\ \sigma_z &\longmapsto z\sigma_z\end{aligned}$$

where x, y, and z are some fixed real numbers.

Using the Choi matrix of $\mathcal{E}$, determine the values of x, y, and z for which the map $\mathcal{E}$ is positive, and the range for which it is completely positive.

10.5.6 Block matrices and the partial trace

For any matrix M in $\mathcal{H}_\mathcal{A} \otimes \mathcal{H}_\mathcal{B}$ *that is written in the tensor product basis*, the partial trace over the first subsystem (here $\mathcal{A}$) is the sum of the diagonal block matrices, and the partial trace over the second subsystem (here $\mathcal{B}$) is a matrix in which the block sub-matrices are replaced by their traces. You can visualise this as in Figure 42.



$\operatorname{tr}_\mathcal{A} M$ $\qquad$ $\operatorname{tr}_\mathcal{B} M$

Figure 42. Visualising the two partial traces of a matrix *written in the tensor product basis*.

For example, for any M in the tensor product space associated with two qubits, written in the standard basis $\{|00\rangle, |01\rangle, |10\rangle, |11\rangle\}$ in block form as

$$M = \left[\begin{array}{c|c} P & Q \\ \hline R & S \end{array}\right]$$

where P, Q, R, and S are all (2×2) sized sub-matrices, we have that

$$\begin{aligned}\operatorname{tr}_\mathcal{A} M &= P + S\\ \operatorname{tr}_\mathcal{B} M &= \left[\begin{array}{c|c} \operatorname{tr} P & \operatorname{tr} Q \\ \hline \operatorname{tr} R & \operatorname{tr} S \end{array}\right]\end{aligned}$$

The same holds for general M in any $\mathcal{H}_\mathcal{A} \otimes \mathcal{H}_\mathcal{B}$ with such a block form (i.e. $m \times m$ blocks of $(n \times n)$ sized sub-matrices, where $m = \dim \mathcal{H}_\mathcal{A}$ and $n = \dim \mathcal{H}_\mathcal{B}$).

10.5.7 Partial inner product

The tensor product structure brings with it the possibility to do "partial things", such as partial traces and partial inner products. Given $\mathcal{H}_\mathcal{A} \otimes \mathcal{H}_\mathcal{B}$, any vector $|x\rangle \in \mathcal{H}_\mathcal{A}$ defines an anti-linear map $\mathcal{H}_\mathcal{A} \otimes \mathcal{H}_\mathcal{B} \to \mathcal{H}_\mathcal{B}$ called the **partial inner product with $|x\rangle$.** It is first defined on the product vectors $|a\rangle \otimes |b\rangle$ by the formula

$$|a\rangle \otimes |b\rangle \longmapsto \langle x|a\rangle|b\rangle$$

and then extended to other vectors in $\mathcal{H}_\mathcal{A} \otimes \mathcal{H}_\mathcal{B}$ by linearity. Similarly, any $|y\rangle \in \mathcal{H}_\mathcal{B}$ defines a map $\mathcal{H}_\mathcal{A} \otimes \mathcal{H}_\mathcal{B} \to \mathcal{H}_\mathcal{A}$ via

$$|a\rangle \otimes |b\rangle \longmapsto |a\rangle\langle y|b\rangle$$

For example, the partial inner product of

$$|\psi\rangle = c_{00}|00\rangle + c_{01}|01\rangle + c_{10}|10\rangle + c_{11}|11\rangle \in \mathcal{H}_\mathcal{A} \otimes \mathcal{H}_\mathcal{B}$$

with of $|0\rangle \in \mathcal{H}_\mathcal{A}$ is

$$\langle 0|\psi\rangle = c_{00}|0\rangle + c_{01}|1\rangle$$

and the partial inner product of the same $|\psi\rangle$ with $|1\rangle \in \mathcal{H}_\mathcal{B}$ is

$$\langle 1|\psi\rangle = c_{01}|0\rangle + c_{11}|1\rangle.$$

10.5.8

Show that $\mathcal{E}$ is trace preserving if and only if $\mathrm{tr}_\mathcal{B}\, \widetilde{\mathcal{E}} = d\mathbf{1}$. Here the partial trace is over the second system in our definition of $\widetilde{\mathcal{E}}$ as $(\mathcal{I} \otimes \mathcal{E})|\Omega\rangle\langle\Omega|$.

Hint. Show that $\mathrm{tr}[\mathcal{E}(|i\rangle\langle j|)] = \delta_{ij}$, and then consider Figure 42.

10.5.9

Mathematically speaking, Kraus operators E_k are vectors in a (dd')-dimensional Hilbert space, with the Hilbert–Schmidt inner product given by $\mathrm{tr}\, E_k^\dagger E_l$. Thus, our intuition tells us, in order to describe any quantum channel with $\dim \mathcal{H} = \dim \mathcal{H}' = d$, we should need no more than d^2 Kraus operators. We can pick any orthonormal basis of operators $\{B_i\}$ and express each Kraus vector in this basis as $E_k = \sum c_{ki} B_i$ (where $i = 1, \ldots, d^2$, $k = 1, \ldots, n$, and n can be much larger than d^2). This gives us

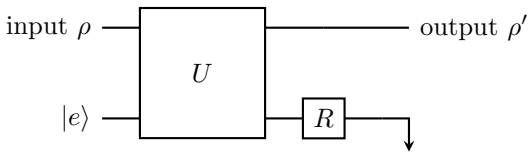
$$\begin{aligned}\rho &\longmapsto \sum_{ij} B_i \rho B_j^\dagger \left(\sum_k c_{ki} c_{kj}^\star \right) \\ &= \sum_{ij} B_i \rho B_j^\dagger C_{ij}\end{aligned}$$

where the matrix C_{ij} is positive semidefinite, and hence unitarily diagonalisable: $C_{ij} = \sum_k U_{ik} d_k U_{kj}^\dagger$ for some unitary U and some real $d_k \geqslant 0$. We can then unitarily "rotate" our operator basis, and use $C_k = \sum_j U_{jk} B_j \sqrt{d_k}$ as our new Krauss operators.

10.5.10 Operator sum freedom

Another way to understand the freedom in the Kraus representation is to realise that any unitary operation *on the environment alone* does not affect the principal system, i.e. the

quantum channel $\mathcal{E}$ is not affected by the choice of the unitary R in the diagram below.



This is true, even though the Kraus operators $E_k = \langle e_k|U|e\rangle$ have now changed to

$$\begin{aligned} F_k &= \langle e_k|(\mathbf{1}\otimes R)U|e\rangle \\ &= \sum_j \langle e_k|R|e_j\rangle\langle e_j|U|e\rangle \\ &= \sum_j R_{kj}E_j \end{aligned}$$

Indeed, the unitary evolution $(\mathbf{1}\otimes R)U$ gives

$$\rho\otimes|e\rangle\langle e| \longmapsto \sum_{kl} E_k\rho E_l^\dagger \otimes R|e_k\rangle\langle e_l|R^\dagger,$$

and the subsequent trace over the environment gives

$$\begin{aligned} \mathrm{tr}_E \sum_{kl} E_k\rho E_l^\dagger \otimes R|e_k\rangle\langle e_l|R^\dagger &= \sum_{kl} E_k\rho E_l^\dagger \langle e_l|R^\dagger R|e_k\rangle \\ &= \sum_k E_k\rho E_k^\dagger. \end{aligned}$$

10.5.11

Consider two single-qubit channels

$$\begin{aligned} \mathcal{E}\colon \rho &\longmapsto \sum_k E_k\rho E_k^\dagger \\ \mathcal{F}\colon \rho &\longmapsto \sum_k F_k\rho F_k^\dagger \end{aligned}$$

defined by their respective Kraus operators

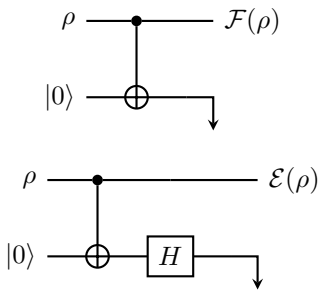
$$\begin{aligned} E_1 &= \frac{1}{\sqrt{2}}\mathbf{1} = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \\ E_2 &= \frac{1}{\sqrt{2}}Z = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \\ F_1 &= |0\rangle\langle 0| = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \\ F_2 &= |1\rangle\langle 1| = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}. \end{aligned}$$

The first channel chooses randomly, with equal probability, between the two options: it will either let the qubit pass undisturbed, or apply the phase flip. The second channel essentially performs the measurement in the standard basis, but the outcome of the measurement is not revealed. These two apparently very different physical processes correspond to the *same* quantum channel: $\mathcal{E}(\rho) = \mathcal{F}(\rho)$ for any ρ.

Indeed, you can check that $E_1 = (F_1 + F_2)/\sqrt{2}$ and $E_2 = (F_1 - F_2)/\sqrt{2}$, whence

$$\begin{aligned} E(\rho) &= E_1 \rho E_1^\dagger + E_2 \rho E_2^\dagger \\ &= \frac{1}{2}(F_1 + F_2)\rho(F_1 + F_2)^\dagger + \frac{1}{2}(F_1 - F_2)\rho(F_1 - F_2)^\dagger \\ &= F_1 \rho F_1^\dagger + F_2 \rho F_2^\dagger \\ &= F(\rho). \end{aligned}$$

You can also check that the two channels can be implemented by the following two circuits:



The `c-NOT` gate appears here as the measurement gate. The target qubit measures the control qubit in either the standard basis (the operation $\mathcal{F}$) or in the Hadamard basis (the operation $\mathcal{E}$). The extra Hadamard gate on the target qubit has no effect on the control qubit.

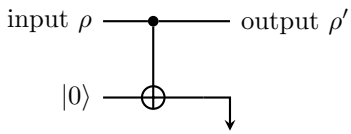
10.5.12 C-NOT, again

We can study the above example more generally, as follows. Let us consider a single-qubit channel induced by the action of the `c-NOT` gate. Recall that the unitary operator associated with the `c-NOT` gate can be written as

$$U = |0\rangle\langle 0| \otimes \mathbf{1} + |1\rangle\langle 1| \otimes X$$

where is X is the Pauli operator σ_x, i.e. the `NOT` gate.

Let us step through the following simple circuit:



The control qubit will be our system, and the target qubit, initially in a fixed state $|e\rangle \equiv |0\rangle$, will play the role of the environment. Let the $\{|e_1\rangle, |e_2\rangle\}$ basis be the standard basis of the target qubit $\{|0\rangle, |1\rangle\}$. In this case we have

Be careful here, the proliferation of $|0\rangle$ and $|1\rangle$ can be confusing. Make sure you do indeed apply $\langle 0|\square|0\rangle$ and $\langle 1|\square|0\rangle$ *only to the environment*, i.e. to the second term in the tensor products.

$$\begin{aligned} E_1 &= \langle e_1|U|e\rangle \\ &= \langle 0|U|0\rangle \\ &= \langle 0|\big(|0\rangle\langle 0| \otimes \mathbf{1} + |1\rangle\langle 1| \otimes X\big)|0\rangle \\ &= |0\rangle\langle 0|\langle 0|\mathbf{1}|0\rangle + |1\rangle\langle 1|\langle 0|X|0\rangle \\ &= |0\rangle\langle 0| \end{aligned}$$

and

$$\begin{aligned} E_2 &= \langle e_2|U|e\rangle \\ &= \langle 1|U|0\rangle \\ &= \langle 1|\big(|0\rangle\langle 0|\otimes \mathbf{1} + |1\rangle\langle 1|\otimes X\big)|0\rangle \\ &= |0\rangle\langle 0|\langle 1|\mathbf{1}|0\rangle + |1\rangle\langle 1|\langle 1|X|0\rangle \\ &= |1\rangle\langle 1| \end{aligned}$$

It is then easy to see that $E_1^\dagger E_1 + E_2^\dagger E_2 = |0\rangle\langle 0| + |1\rangle\langle 1| = \mathbf{1}$.

The unitary part of the evolution

$$\begin{aligned} |\psi\rangle|e\rangle \longmapsto & E_1|\psi\rangle|e_1\rangle + E_2|\psi\rangle|e_2\rangle \\ = & |0\rangle\langle 0|\psi\rangle|0\rangle + |1\rangle\langle 1|\psi\rangle|1\rangle \\ = & \langle 0|\psi\rangle|0\rangle|0\rangle + \langle 1|\psi\rangle|1\rangle|1\rangle \end{aligned}$$

is a familiar `c-NOT` entangling process: if $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$, then $|\psi\rangle|0\rangle$ becomes $\alpha|0\rangle|0\rangle + \beta|1\rangle|1\rangle$.

The final state of the system qubit, the output, is given by

$$\begin{aligned} \rho' &= E_1\rho E_1^\dagger + E_2\rho E_2^\dagger \\ &= |0\rangle\langle 0|\rho|0\rangle\langle 0| + |1\rangle\langle 1|\rho|1\rangle\langle 1| \\ &= \rho_{00}|0\rangle\langle 0| + \rho_{11}|1\rangle\langle 1| \end{aligned}$$

or, in matrix form, by

$$\begin{bmatrix} |\alpha|^2 & \alpha\beta^\star \\ \alpha^\star\beta & |\beta|^2 \end{bmatrix} = \rho \longmapsto \rho' = \begin{bmatrix} |\alpha|^2 & 0 \\ 0 & |\beta|^2 \end{bmatrix}.$$

As we can see, the diagonal elements of ρ survive, and the off-diagonal elements (i.e. the coherences) disappear. You may think about this operation as being equivalent to measuring the system qubit in the standard basis and then forgetting the result.

11 Quantum error correction and fault tolerance