

Máté Farkas

QUANTUM THEORY & QUANTUM INFORMATION


Lecture Notes – Autumn 2023

These notes are based on Matthew Pusey's Quantum Mechanics 1 and Quantum Information notes for 2021 and 2023, which were in turn based on notes by Daniel Siemssen and Stefan Weigert.

Contents

Contents	1
1 Introduction	3
2 Postulates of quantum mechanics	6
2.1 States	6
2.2 Measurements	11
2.3 Dynamics	22
3 Qubits	26
3.1 Bloch sphere	26
3.2 BB84 protocol	28
4 Composite systems	32
4.1 Tensor product	32
4.2 State space	35
4.3 Entanglement	38
4.4 Operators	40
4.5 Measurements	43
5 Density operators	45
5.1 Quantum states as operators	45
5.2 Probabilistic mixtures of quantum states	48
5.3 Postulates revisited	51
5.4 Bloch ball	52
5.5 Reduced density operator	53
5.6 Entanglement and no-signalling	56
5.7 Purification of states	58
6 Quantum circuits	60
6.1 No-cloning theorem	60
6.2 Quantum gates and circuit diagrams	61
6.3 Quantum teleportation	64
6.4 Decomposition of unitary operators	66
6.5 Universal quantum computer	68
7 Quantum algorithms	70
7.1 Quantum parallelism	70

7.2	Deutsch's algorithm	72
7.3	Deutsch–Jozsa algorithm	73
7.4	Subroutine: change phase of subset	74
7.5	Grover's algorithm	75

WARNING. These notes include a few results and remarks which go beyond the scope of this module. I have marked them with a  (‘dangerous bend’) symbol in the left margin. They are not examinable and can thus safely be skipped.

1 Introduction

Quantum theory—like every physical theory—aims to describe the physical world around us. It provides a mathematical *description* of physical systems, and also makes *predictions* on how these systems will behave. Importantly, these descriptions and predictions can be verified by (experimental) *observation*.

Quantum theory was mostly developed in the early twentieth century, after difficulties arose in describing certain physical systems and processes—such as the hydrogen atom or black-body radiation—by classical physics (a.k.a. Newtonian mechanics). Since then, quantum theory has been the most successful theory in describing subatomic particles, and its predictions have been experimentally verified with unprecedented precision.

While quantum theory is extremely useful and has so far been successful in describing everything we see on the subatomic level, there is one somewhat unsettling aspect to it that is quite different from other physical theories. Newtonian mechanics, and even Einstein’s theory of relativity, is based on *physical principles*. On the other hand, quantum theory is an *axiomatic* theory: it is entirely based on mathematical axioms. Of course, Newtonian mechanics or special relativity can also be (and in the working practice, mostly is) described mathematically, but the mathematical laws have clear physical principles behind them. For example, Newton’s second law is given by

$$\vec{F} = m\vec{a},$$

mathematically speaking. But there is an equivalent physical principle that is easy to articulate: *The change of motion of an object is proportional to the force impressed; and is made in the direction of the straight line in which the force is impressed.*¹ Similarly, special relativity relies on two physical principles: 1) *The laws of physics are the same for all observers in any inertial frame of reference relative to one another.* 2) *The speed of light in a vacuum is the same for all observers, regardless of their relative motion or of the motion of the light source.*² These principles are intuitive (at least to physicists), and so physicists can sleep well at night knowing that their theories are based on these well-founded principles.

Quantum theory is vastly different in this regard. It is entirely based on mathematical *axioms*, also called *the postulates of quantum theory*. For example, the first postulate reads (see also later in Chapter 2):

A quantum mechanical system is associated with a complex Hilbert space \mathcal{H} . The state of the system is represented by a normalised element of

¹ This particular phrasing of Newton’s second law can be found on Wikipedia.

² Again, Wikipedia.

the space \mathcal{H} . Two normalised elements $\varphi, \psi \in \mathcal{H}$ represent the same state if $\varphi = \alpha\psi$ for some complex number $\alpha \in \mathbb{C}$.

And this is as intuitive as it gets. It is a whole (and currently active) line of research to re-derive quantum theory from some physical principles instead of axioms of these sorts. In this process (and the scientific process of trying to come up with better and better theories), quantum theory might need to be amended or completely replaced, or maybe this research programme will be successful and we will have a formulation of quantum theory that is more intuitive and easier to explain to the layman (and physicists). But until then, we are stuck with the postulates of quantum theory.

Fortunately, the axiomatic nature of quantum theory makes it well-suited to introduce it to mathematicians. Indeed, one doesn't need to care about the “intuitive meaning” of these axioms (since there are barely any!), and can treat them as a handy list of axioms from which to develop a theory. *Quantum information theory* often takes this abstraction one level further, and is only concerned with describing the *probabilities* of certain types of events happening. These events are often abstract enough (e.g. “if I press a button on my measurement device, what is the probability that I observe outcome A or outcome B?”) that the description is not sensitive to the actual physical implementation. Maybe the experiment is done using photons, maybe electrons, perhaps even *tardigrades*³—as long as the system can be described by quantum theory, the quantum information theoretical description is valid. As such, students attending this module (and people interested in quantum information theory, in general) don't need to worry about knowing much (or anything) about physics.

The main aim of quantum information theory is to investigate how to use quantum mechanical systems for information processing (e.g. computation or communication). It turns out that information processing protocols based on quantum theory sometimes have an (for now, theoretical) advantage over “classical” protocols (the ones we use today, based on classical physics). For example, certain computational problems that are believed to be hard to solve classically, can be solved efficiently on large-scale quantum hardware. Also, certain communication tasks require less quantum resources than equivalent classical resources. Furthermore, protocols based on quantum theory provide unconditional cryptographic security, based on the laws of nature rather than on computational hardness, the assumption used in current classical protocols. These advantages in information processing are often related to fundamental questions in quantum theory. The quantum advantage highlights the difference between classical and quantum theory, and often helps pinpointing particular ways in which they differ. Moreover, the *extent* of the advantage achievable motivates us to explore the fundamental limitations of quantum theory.

In this module, we will first go through the postulates of quantum theory in a finite-dimensional setting: no hydrogen atom or wave-function of an electron, only what is essential to understand most things in quantum information theory. We will cover the basics of Hilbert spaces, quantum states and measurements, how to obtain measurement outcome probabilities, how

³ This is an in-joke and shouldn't be taken too seriously. A complete quantum mechanical description of a tardigrade is way too complicated to carry out in any meaningful way. For reference, see e.g. [this article](#).

quantum systems evolve and what is quantum entanglement. We will put this knowledge into a quantum information context, and discuss fundamental and practical concepts, such as no-signalling, no-cloning, teleportation, quantum cryptography, quantum computing, quantum algorithms and Bell non-locality.

2 Postulates of quantum mechanics

2.1 STATES

The *state* of a physical system (an apple, a spaceship, an electron, ...) refers to its full (mathematical) description in a given moment of time. In classical (non-quantum) physics, systems are generally described in *phase space*: A system is characterised by its *position* (where the system is) and *momentum* (roughly speaking, “in which direction and how fast the object is moving”), and derived quantities.

In quantum mechanics, position and momentum are a bit more tricky, and in general one can only assign probabilities to a given system’s position and momentum. As such, the mathematical description of quantum states is more difficult to grasp intuitively, and the theory is entirely built on mathematical axioms: the postulates of quantum theory. The first postulate tells us what are the mathematical objects that we can use to describe the state of a physical system:

POSTULATE (STATES). A (quantum mechanical) system is associated with a complex Hilbert space \mathcal{H} . The *state* of the system is represented by a normalised element of the space \mathcal{H} . Two normalised elements $\varphi, \psi \in \mathcal{H}$ represent the same state if $\varphi = \alpha\psi$ for some complex number $\alpha \in \mathbb{C}$. \diamond

In order to fully understand this postulate, we begin by reviewing the notions necessary to define a Hilbert space. For concreteness and brevity, we will restrict our attention to the complex setting—the only setting which is relevant for quantum mechanics. Note, however, that all concepts defined here carry over to the real setting with very few modifications.

2.1.1 Hilbert spaces

A Hilbert space is a special case of a *vector space*, and therefore we first revisit the definition of a vector space.

DEFINITION 2.1. A *vector space* (over \mathbb{C}) is a triple $(\mathcal{V}, +, \cdot)$, where

1. \mathcal{V} is a set,
2. $+: \mathcal{V} \times \mathcal{V} \rightarrow \mathcal{V}$, $(u, v) \mapsto u + v$ (addition),
3. $\cdot: \mathbb{C} \times \mathcal{V} \rightarrow \mathcal{V}$, $(\alpha, u) \mapsto \alpha \cdot u$ (scalar multiplication)

satisfying:

1. $u + v = v + u$ for all $u, v \in \mathcal{V}$ (commutativity of $+$),
2. $(u + v) + w = u + (v + w)$ for all $u, v, w \in \mathcal{V}$ (associativity of $+$),
3. there exists $\vec{0} \in \mathcal{V}$ such that $\vec{0} + v = v$ for all $v \in \mathcal{V}$ (additive identity),
4. given any $v \in \mathcal{V}$ there exists $(-v) \in \mathcal{V}$ with $(-v) + v = \vec{0}$ (additive inverse),
5. $\alpha \cdot (\beta \cdot v) = (\alpha\beta) \cdot v$ for all $\alpha, \beta \in \mathbb{C}$ and $v \in \mathcal{V}$ (compatibility of scalar multiplication and multiplication in \mathbb{C}),
6. $\alpha \cdot (u + v) = \alpha \cdot u + \alpha \cdot v$ for all $\alpha \in \mathbb{C}$ and $u, v \in \mathcal{V}$ (distributivity of scalar multiplication with respect to vector addition),
7. $(\alpha + \beta) \cdot v = \alpha \cdot v + \beta \cdot v$ for all $\alpha, \beta \in \mathbb{C}$ and $v \in \mathcal{V}$ (distributivity of scalar multiplication with respect to addition in \mathbb{C}),
8. $1 \cdot v = v$ for all $v \in \mathcal{V}$ (identity of scalar multiplication).

TERMINOLOGY & NOTATION.

- We often informally say that “ \mathcal{V} is a vector space”, with the binary operations $+$ and \cdot understood implicitly.
- We usually omit the dot symbol ‘ \cdot ’ when writing scalar multiplications.
- We will (essentially) always write $\vec{0}$ as 0.

EXAMPLE 2.3. Consider the set of complex n -tuples \mathbb{C}^n with the usual rules of addition and scalar multiplication. That is, for $u = (u_1, \dots, u_n) \in \mathbb{C}^n$, $v = (v_1, \dots, v_n) \in \mathbb{C}^n$ and $\alpha \in \mathbb{C}$, define

$$u + v = (u_1 + v_1, \dots, u_n + v_n) \quad \text{and} \quad \alpha \cdot v = (\alpha v_1, \dots, \alpha v_n).$$

Then, $(\mathbb{C}^n, +, \cdot)$ is a vector space (for each n). \diamond

CAUTION. Often we use the same symbols, e.g., $+$ and \cdot , in different contexts. In the example above, $+$ denotes both the addition of tuples and the addition in \mathbb{C} , but there is no ambiguity as to which operation is meant. \diamond

Important concepts in vector spaces are *span*, *linear independence*, *basis* and *dimension*.

DEFINITION 2.5. Consider a (finite) set of vectors $\{e_i\}_{i=1}^n \subset \mathcal{V}$.

1. The set $\{e_i\}_{i=1}^n$ *spans* the space \mathcal{V} if every vector $v \in \mathcal{V}$ can be written as

$$v = \sum_{i=1}^n \alpha_i e_i \tag{2.1}$$

for some complex numbers α_i .

2. The set $\{e_i\}_{i=1}^n$ is *linearly independent* if for a set of complex numbers $\{\alpha_i\}_{i=1}^n$

$$\sum_{i=1}^n \alpha_i e_i = 0$$

implies $\alpha_i = 0$ for all $i \in \{1, \dots, n\}$.

3. If the set $\{e_i\}_{i=1}^n$ spans the space \mathcal{V} and is linearly independent, it is called a *basis* of \mathcal{V} . In this case, n is called the *dimension* of \mathcal{V} , denoted $\dim \mathcal{V}$, and every basis of \mathcal{V} has n elements. Furthermore, for every $v \in \mathcal{V}$, the decomposition in Eq. (2.1) is unique. \diamond

REMARK 2.6. In this module, we will only consider finite-dimensional vector spaces (unless explicitly stated otherwise, in some examples that only serve demonstration purposes). Most definitions carry through to the infinite-dimensional case, but these often need extra care and a few more technical assumptions. We will not need to worry about these technicalities in this module. \diamond

In order for a vector space to be a Hilbert space, some additional structure is necessary. In particular, the vector space must be equipped with an *inner product*.

DEFINITION 2.7. A (*Hermitian*) *inner product* on a vector space \mathcal{V} is a map

$$\langle \cdot, \cdot \rangle : \mathcal{V} \times \mathcal{V} \rightarrow \mathbb{C}$$

satisfying

1. $\overline{\langle v, u \rangle} = \langle u, v \rangle$ for all $u, v \in \mathcal{V}$, where $\overline{(\cdot)}$ is the complex conjugation,
2. $\langle u, v + w \rangle = \langle u, v \rangle + \langle u, w \rangle$ for all $u, v, w \in \mathcal{V}$,
3. $\langle u, \alpha v \rangle = \alpha \langle u, v \rangle$ for all $u, v \in \mathcal{V}$ and $\alpha \in \mathbb{C}$,
4. $\langle v, v \rangle \geq 0$ for all $v \in \mathcal{V}$ with equality if and only if $v = 0$.

A vector space with an inner product is called a (*Hermitian*) *inner product space*. We denote by $\|\cdot\|$ the *norm* on \mathcal{V} defined by $\|v\| := \sqrt{\langle v, v \rangle}$ for all $v \in \mathcal{V}$. \diamond

EXAMPLE 2.8. We define on \mathbb{C}^n the usual Euclidean inner product by

$$\langle u, v \rangle = u^\dagger v = \bar{u}^T v,$$

where the complex conjugation and the transpose $(\cdot)^T$ are in the same (but arbitrary) basis. \diamond

TERMINOLOGY & NOTATION. Throughout these notes, we will use the notation $(\cdot)^\dagger$ for the *Hermitian conjugate* or *Hermitian adjoint*, that is, for the conjugate transpose. Some references (especially in algebra) might use the notation $(\cdot)^*$. At the same time, some other references often use the notation $(\cdot)^*$ for the complex conjugate. To avoid confusion, we will not use the notation $(\cdot)^*$ in this module. \diamond

DEFINITION 2.10. Let \mathcal{V} be an inner product space.

1. $v \in \mathcal{V}$ is called *normalised* if $\langle v, v \rangle = \|v\|^2 = 1$.
2. $u, v \in \mathcal{V}$ are called *orthogonal* if $\langle u, v \rangle = 0$.
3. $u, v \in \mathcal{V}$ are called *orthonormal* if they are orthogonal and normalised.
4. A basis $\{e_i\}_{i=1}^n \subset \mathcal{V}$ is called *orthonormal (ONB)* if its elements are normalised and pair-wise orthogonal. \diamond

THEOREM 2.11. Let \mathcal{V} be an inner product space. For all $u, v \in \mathcal{V}$ we have the *Cauchy–Schwarz inequality*

$$|\langle u, v \rangle| \leq \|u\| \|v\|$$

and the *triangle inequality*

$$||| \|u\| - \|v\| \leq \|u + v\| \leq \|u\| + \|v\|.$$

In the finite-dimensional case, the definition of a Hilbert space is equivalent to the definition of an inner product space:

DEFINITION 2.12. A finite-dimensional *Hilbert space*, often denoted by \mathcal{H} , is a finite-dimensional inner product space. \diamond

EXAMPLE 2.13. \mathbb{C}^n with the Euclidean inner product is a Hilbert space. \diamond

This is actually the prototypical finite-dimensional Hilbert space as the following result shows:

THEOREM 2.14. Every n -dimensional inner product space \mathcal{V} is ‘equivalent’ (isometric) to \mathbb{C}^n with the Euclidean inner product.

PROOF. Choose an orthonormal basis $\{e_1, \dots, e_n\} \subset \mathcal{V}$. Since every vector $v \in \mathcal{V}$ can be uniquely decomposed as

$$v = \alpha_1 e_1 + \dots + \alpha_n e_n, \quad \alpha_1, \dots, \alpha_n \in \mathbb{C},$$

we can identify it with the n -tuple $(\alpha_1, \dots, \alpha_n) \in \mathbb{C}^n$. This shows that \mathcal{V} and \mathbb{C}^n are isomorphic as vector spaces.

Now suppose that $u, v \in \mathcal{V}$ are represented by the vectors $(\alpha_1, \dots, \alpha_n)$ and $(\beta_1, \dots, \beta_n)$. Then their inner product is

$$\langle u, v \rangle = \sum_{i=1}^n \sum_{j=1}^n \langle \alpha_i e_i, \beta_j e_j \rangle = \sum_{i=1}^n \sum_{j=1}^n \overline{\alpha_i} \beta_j \langle e_i, e_j \rangle = \sum_{i=1}^n \overline{\alpha_i} \beta_i,$$

which is just the Euclidean product of the vectors $(\alpha_1, \dots, \alpha_n)$ and $(\beta_1, \dots, \beta_n)$. This shows that \mathcal{V} and \mathbb{C}^n are isometric as Hilbert spaces. \square

As a consequence of the above theorem, every Hilbert space in this module (unless explicitly stated otherwise) can be thought of the Hilbert space \mathbb{C}^n for some positive integer n . According to the state postulate, every state is then a normalised element of \mathbb{C}^n , that is, an element

$$v = (v_1, \dots, v_n) \in \mathbb{C}^n,$$

such that

$$\|v\| = \sqrt{\langle v, v \rangle} = \sqrt{|v_1|^2 + \dots + |v_n|^2} = 1.$$

EXAMPLE 2.15. Consider the Hilbert space \mathbb{C}^2 and an orthonormal basis $\{e_1, e_2\}$ on it. Then every state can be written as

$$\psi = \alpha e_1 + \beta e_2,$$

where $\alpha, \beta \in \mathbb{C}$ and $|\alpha|^2 + |\beta|^2 = 1$. Furthermore, for every complex number of modulus one, $e^{i\theta}$ where $\theta \in \mathbb{R}$, the state

$$\varphi = e^{i\theta} \psi = e^{i\theta} \alpha e_1 + e^{i\theta} \beta e_2$$

represents the same state as ψ (also referred to as *equivalent*). ◇

2.1.2 Infinite-dimensional Hilbert spaces

While infinite-dimensional Hilbert spaces are beyond the scope of this module, we will review a few basic concepts and examples in this section that might make it easier to draw connections between this module and the quantum parts of the Classical & Quantum Dynamics module. **For the purpose of revising, you may freely skip this section.**

In the infinite-dimensional case, the definition of inner product spaces is not equivalent to the definition of Hilbert spaces. In general, a Hilbert space is a *complete* inner product space.

DEFINITION 2.16. A vector space \mathcal{V} with a norm $\|\cdot\|$ is called *complete* if for every sequence $(v_n)_{n \in \mathbb{N}}$, $v_n \in \mathcal{V}$, such that

$$\lim_{n, m \rightarrow \infty} \|v_n - v_m\| = 0$$

there exists a $v \in \mathcal{V}$, called the limit of the sequence $(v_n)_{n \in \mathbb{N}}$, such that

$$\lim_{n \rightarrow \infty} \|v - v_n\| = 0$$

(in other words, every Cauchy sequence converges). ◇

DEFINITION 2.17. A *Hilbert space* is a complete inner product space. ◇

EXAMPLE 2.18. Consider the set of square-integrable complex-valued functions on \mathbb{R}^n ,

$$L^2(\mathbb{R}^n) = \left\{ f : \mathbb{R}^n \rightarrow \mathbb{C} : \int_{\mathbb{R}^n} |f(\vec{x})|^2 d\vec{x} < \infty \right\}.$$

We define addition and scalar multiplication of functions point-wise:

1. For $f, g \in L^2(\mathbb{R}^n)$, define $f + g$ by $(f + g)(\vec{x}) = f(\vec{x}) + g(\vec{x})$ for all $\vec{x} \in \mathbb{R}^n$,
2. For $f \in L^2(\mathbb{R}^n)$ and $\alpha \in \mathbb{C}$, define $\alpha \cdot f$ by $(\alpha \cdot f)(\vec{x}) = \alpha f(\vec{x})$ for all $\vec{x} \in \mathbb{R}^n$,

$(L^2(\mathbb{R}^n), +, \cdot)$ is a vector space (for each n). To make it a Hilbert space, we define an inner product

$$\langle f, g \rangle = \int_{\mathbb{R}^n} \overline{f(\vec{x})} g(\vec{x}) \, d\vec{x}, \quad f, g \in L^2(\mathbb{R}^n). \quad (2.2)$$

It can be proven that $L^2(\mathbb{R}^n)$ with the norm induced by the above inner product is complete, and therefore a Hilbert space. \diamond

REMARK 2.19. If you look closely, you will notice that many non-negative functions have a vanishing integral. For this reason it seems that the inner product defined above is not actually positive definite! We can circumvent this problem by identifying all non-negative functions that have a vanishing integral with the zero function, that is, one should actually look at equivalence classes of functions and change the definition of $L^2(\mathbb{R}^n)$ above. \diamond

The above example might be familiar from the Classical & Quantum Dynamics module, where wave functions of physical systems living in three spatial dimensions are defined by functions $\psi : \mathbb{R}^{3+1} \rightarrow \mathbb{C}$. In particular, ψ maps a three-dimensional coordinate \vec{x} and a time parameter t to $\psi(\vec{x}, t)$. The wave function must be *normalised*, that is,

$$\int_{\mathbb{R}^3} |\psi(\vec{x}, t)|^2 \, d\vec{x} = 1.$$

One can see that for every value of t , the wave function is a normalised element of the Hilbert space $L^2(\mathbb{R}^3)$ (remember: a quantum state is the mathematical description of a physical system in a given moment of time). In this module, states are still associated with a normalised element of a Hilbert space, but the Hilbert space will always be finite-dimensional.

2.2 MEASUREMENTS

Measurements extract *information* about a physical system. That is, a measurement takes a state as an input and produces a *label* (the measurement outcome) as an output, for example *the particle is at position x* . In classical physics, the position and momentum of a physical system fully characterise it, and therefore measuring its position and momentum suffices to give us complete knowledge about the state. In quantum physics, the position and momentum of a particle cannot be measured at the same time with arbitrary precision. This phenomenon is ubiquitous in quantum theory, and in general we can only assign a *probability* to a given measurement outcome.

If we measure a quantum particle, we will observe a particular outcome. However, repeating this exact same experiment might result in a different

outcome the next time. Therefore, to give a full characterisation of a quantum measurement, we need to define an object that takes a state as an input, and as an output, it gives us the probabilities with which we will observe each of its possible outcomes, and also the state of the system after obtaining a particular measurement outcome.

For a given state on a Hilbert space \mathcal{H} , measurements are associated with certain types of operators on \mathcal{H} . In order to be able to state the measurement postulate, we first review some properties of operators on Hilbert spaces.

2.2.1 Operators on Hilbert spaces

DEFINITION 2.20. Let \mathcal{H}, \mathcal{K} be Hilbert spaces and $A : \mathcal{H} \rightarrow \mathcal{K}$ a map. Then A is called a (*linear*) *operator* (also linear transformation or linear map) from \mathcal{H} to \mathcal{K} if the following holds:

1. $A(u + v) = A(u) + A(v)$ for all $u, v \in \mathcal{H}$,
2. $A(\alpha v) = \alpha A(v)$ for all $v \in \mathcal{H}$ and $\alpha \in \mathbb{C}$.

We denote the set of linear operators from \mathcal{H} to \mathcal{K} by $\mathcal{L}(\mathcal{H}, \mathcal{K})$. If $\mathcal{H} = \mathcal{K}$, we say that A is a linear operator on \mathcal{H} , and we denote the set of linear operators on \mathcal{H} by $\mathcal{L}(\mathcal{H})$. \diamond

REMARK 2.21. Note that $\mathcal{L}(\mathcal{H}, \mathcal{K})$ is itself a vector space. If $A, B \in \mathcal{L}(\mathcal{H}, \mathcal{K})$ then one can define $(A + B) \in \mathcal{L}(\mathcal{H}, \mathcal{K})$ via its action

$$(A + B)(v) = A(v) + B(v)$$

for all $v \in \mathcal{H}$. Similarly, one can define $\alpha A \in \mathcal{L}(\mathcal{H}, \mathcal{K})$ for an arbitrary $\alpha \in \mathbb{C}$ via its action

$$(\alpha A)(v) = \alpha A(v)$$

for all $v \in \mathcal{H}$. It is straightforward to verify that $\mathcal{L}(\mathcal{H}, \mathcal{K})$ with these operations is indeed a vector space. \diamond

TERMINOLOGY & NOTATION.

- When applying a linear operator to a single symbol, we often omit the parentheses and write, e.g. Av instead of $A(v)$.
- In this module we will only consider *linear* operators (as opposed to *non-linear* operators) and thus will drop the qualifier “linear”. \diamond

PROPOSITION 2.23. If A is an operator from \mathcal{H} to \mathcal{K} , then A is fully characterised by, equivalently,

1. its action Av on all $v \in \mathcal{H}$
2. its action Ae_i on a basis $\{e_i\}$ of \mathcal{H}
3. the inner products $\langle v, Aw \rangle$ for all $w \in \mathcal{H}$ and $v \in \mathcal{K}$

4. the inner products $\langle e_i, Af_j \rangle$, where $\{f_j\}$ is a basis of \mathcal{H} and $\{e_i\}$ is a basis of \mathcal{K}

EXAMPLE 2.24. Let \mathcal{H} be an arbitrary Hilbert space. The *identity map* $\mathbb{1}$, defined by $\mathbb{1}v = v$ for all $v \in \mathcal{V}$, is an operator. \diamond

EXAMPLE 2.25. Every matrix $A \in M_{m \times n}(\mathbb{C})$ defines an operator from \mathbb{C}^n to \mathbb{C}^m by multiplication. In fact, every operator from \mathbb{C}^n to \mathbb{C}^m is given by a matrix in $M_{m \times n}(\mathbb{C})$. \diamond

A crucial concept in Hilbert space theory—and especially in quantum theory—is *adjoint operators*.

DEFINITION 2.26. Let \mathcal{H} and \mathcal{K} be Hilbert spaces. For any operator $A : \mathcal{H} \rightarrow \mathcal{K}$ there is a unique operator $A^\dagger : \mathcal{K} \rightarrow \mathcal{H}$, called the *adjoint* of A , such that

$$\langle \varphi, A\psi \rangle_{\mathcal{K}} = \langle A^\dagger \varphi, \psi \rangle_{\mathcal{H}} \quad \text{or, equivalently,} \quad \langle \psi, A^\dagger \varphi \rangle_{\mathcal{H}} = \langle A\psi, \varphi \rangle_{\mathcal{K}} \quad (2.3)$$

for all $\psi \in \mathcal{H}$ and $\varphi \in \mathcal{K}$, where $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ is the inner product on \mathcal{H} and $\langle \cdot, \cdot \rangle_{\mathcal{K}}$ is the inner product on \mathcal{K} . \diamond

The following example justifies the notation $(\cdot)^\dagger$ that we used earlier for the conjugate transpose.

EXAMPLE 2.27. Consider the Hilbert spaces \mathbb{C}^n and \mathbb{C}^m (with the Euclidean inner product). Recall that every operator from \mathbb{C}^n to \mathbb{C}^m is given by a matrix $A \in M_{m \times n}(\mathbb{C})$. For any $u \in \mathbb{C}^m$ and $v \in \mathbb{C}^n$ we have

$$\langle u, Av \rangle_{\mathbb{C}^m} = \bar{u}^T Av = (A^T \bar{u})^T v = \overline{(\bar{A}^T u)}^T v = \langle \bar{A}^T u, v \rangle_{\mathbb{C}^n}.$$

Therefore the adjoint is just the conjugate transpose, that is, $A^\dagger = \bar{A}^T$. \diamond

PROPOSITION 2.28. Let $A, B \in \mathcal{L}(\mathcal{H}, \mathcal{K})$ and $\alpha \in \mathbb{C}$. We have the properties:

1. $(\alpha A)^\dagger = \bar{\alpha} A^\dagger$,
2. $(A + B)^\dagger = A^\dagger + B^\dagger$,
3. $(AB)^\dagger = B^\dagger A^\dagger$ (in the case of $\mathcal{K} = \mathcal{H}$),
4. $(A^\dagger)^\dagger = A$,

PROOF. Let $\psi \in \mathcal{H}$ and $\varphi \in \mathcal{K}$ be arbitrary. The properties follow from the identities:

1. $\langle (\alpha A)^\dagger \varphi, \psi \rangle_{\mathcal{H}} = \langle \varphi, \alpha A\psi \rangle_{\mathcal{K}} = \alpha \langle \varphi, A\psi \rangle_{\mathcal{K}} = \alpha \langle A^\dagger \varphi, \psi \rangle_{\mathcal{H}} = \langle \bar{\alpha} A^\dagger \varphi, \psi \rangle_{\mathcal{H}}$
2. $\langle (A + B)^\dagger \varphi, \psi \rangle_{\mathcal{H}} = \langle \varphi, (A + B)\psi \rangle_{\mathcal{K}} = \langle \varphi, A\psi \rangle_{\mathcal{K}} + \langle \varphi, B\psi \rangle_{\mathcal{K}} \\ = \langle A^\dagger \varphi, \psi \rangle_{\mathcal{H}} + \langle B^\dagger \varphi, \psi \rangle_{\mathcal{H}} = \langle (A^\dagger + B^\dagger) \varphi, \psi \rangle_{\mathcal{H}}$

3. $\langle (AB)^\dagger \varphi, \psi \rangle_{\mathcal{H}} = \langle \varphi, AB\psi \rangle_{\mathcal{H}} = \langle \varphi, A(B\psi) \rangle_{\mathcal{H}} = \langle A^\dagger \varphi, B\psi \rangle_{\mathcal{H}} = \langle B^\dagger A^\dagger \varphi, \psi \rangle_{\mathcal{H}}$
4. $\langle \varphi, (A^\dagger)^\dagger \psi \rangle_{\mathcal{H}} = \langle A^\dagger \varphi, \psi \rangle_{\mathcal{H}} = \langle \varphi, A\psi \rangle_{\mathcal{H}}$ \square

A central concept in quantum theory is that of self-adjoint operators.

DEFINITION 2.29. An operator A on \mathcal{H} is called *self-adjoint* (or *Hermitian*) if $A^\dagger = A$. \diamond

DEFINITION 2.30. Suppose that A is an operator on a Hilbert space \mathcal{H} .

1. The complex number $\lambda \in \mathbb{C}$ is called an *eigenvalue* of A if there exists a vector $\psi \in \mathcal{H}$ such that

$$A\psi = \lambda\psi.$$

We call ψ an *eigenvector* of A , or an *eigenstate* if it is normalised.

2. The set of all eigenvalues of A is called its *spectrum* and we denote it by $\sigma(A)$.
3. An operator A may have more than one linearly independent eigenvector with the same eigenvalue λ . These vectors span a subspace $\mathcal{H}_\lambda \subset \mathcal{H}$ which is called the *eigenspace* corresponding to the eigenvalue λ .
4. The dimension $m_\lambda = \dim \mathcal{H}_\lambda$ of \mathcal{H}_λ is the (*geometric*) *multiplicity* of the eigenvalue λ . If the multiplicity is one, we say that the eigenvalue λ is *non-degenerate*, otherwise it is called *degenerate*. \diamond

THEOREM 2.31. If A is a self-adjoint operator on \mathcal{H} , it holds that

1. the eigenvalues of A are real,
2. $\langle \psi_1, \psi_2 \rangle = 0$, that is, ψ_1 and ψ_2 are orthogonal for all eigenvectors $\psi_1, \psi_2 \in \mathcal{H}$ with distinct eigenvalues.

PROOF.

1. Suppose that $\psi \in \mathcal{H}$ is an eigenvector for A with the eigenvalue $\lambda \in \mathbb{C}$, that is, $A\psi = \lambda\psi$. Since A is self-adjoint, it holds that

$$\lambda \langle \psi, \psi \rangle = \langle \psi, A\psi \rangle = \langle A\psi, \psi \rangle = \bar{\lambda} \langle \psi, \psi \rangle.$$

It follows that $\lambda = \bar{\lambda}$.

2. Let $A\psi_1 = \lambda_1\psi_1$ and $A\psi_2 = \lambda_2\psi_2$ with $\lambda_1 \neq \lambda_2$. Since A is self-adjoint, we find

$$\lambda_1 \langle \psi_1, \psi_2 \rangle = \langle A\psi_1, \psi_2 \rangle = \langle \psi_1, A\psi_2 \rangle = \lambda_2 \langle \psi_1, \psi_2 \rangle.$$

and thus $(\lambda_1 - \lambda_2) \langle \psi_1, \psi_2 \rangle = 0$. But $\lambda_1 - \lambda_2 \neq 0$ and thus $\langle \psi_1, \psi_2 \rangle = 0$ as required. \square

2.2.2 Dirac notation

The commonly used notation for Hilbert spaces in quantum information theory (especially in the physics community) is the *Dirac notation* (or *bra-ket notation*). This is simply a (slightly) different notation for vectors and linear operators in Hilbert space that is often convenient, especially in the context of quantum information theory. While this is just a notation, we will review it in this section to give a solid mathematical justification for its use.

DEFINITION 2.32. Let \mathcal{H} be a Hilbert space. To any vector $v \in \mathcal{H}$, we associate an operator

$$|v\rangle : \mathbb{C} \rightarrow \mathcal{H}, \quad \alpha \mapsto \alpha v,$$

called the *ket* associated to v . \diamond

If $\dim \mathcal{H} = d$ then we are mapping a column vector of length d to a $d \times 1$ matrix with the same elements as those of the column vector (which makes it difficult to see any difference at all!). As such, a ket is just a different way of thinking of a vector. A perhaps more substantial definition is the adjoint of a ket, called a *bra*:

DEFINITION 2.33. Let \mathcal{H} be a Hilbert space. To any vector $v \in \mathcal{H}$, we associate

$$\langle v| = |v\rangle^\dagger,$$

the *bra* associated to v . By definition, $\langle v| : \mathcal{H} \rightarrow \mathbb{C}$. \diamond

The composition of a bra $\langle v| : \mathcal{H} \rightarrow \mathbb{C}$ with a ket $|w\rangle : \mathbb{C} \rightarrow \mathcal{H}$ produces a linear operator $\langle v||w\rangle : \mathbb{C} \rightarrow \mathbb{C}$, which means that $\langle v||w\rangle$ is just multiplication by a complex number (and can thus be identified with an element of \mathbb{C}).

PROPOSITION 2.34. Let \mathcal{H} be a Hilbert space and $v, w \in \mathcal{H}$. Then for all $\alpha \in \mathbb{C}$, the associated bra $\langle v|$ and ket $|w\rangle$ satisfy

$$\langle v||w\rangle(\alpha) = \alpha \langle v, w \rangle_{\mathcal{H}}.$$

PROOF. Recall that if we consider \mathbb{C} as a (one-dimensional) Hilbert space, the inner product of two complex numbers α, β is given by $\langle \alpha, \beta \rangle_{\mathbb{C}} = \bar{\alpha}\beta$. In particular, $\langle 1, \beta \rangle_{\mathbb{C}} = \beta$. Hence

$$\langle v||w\rangle(\alpha) = \langle 1, \langle v||w\rangle(\alpha) \rangle_{\mathbb{C}}.$$

Since $\langle v|^\dagger = |v\rangle$ we have

$$\langle 1, \langle v||w\rangle(\alpha) \rangle_{\mathbb{C}} = \langle |v\rangle(1), |w\rangle(\alpha) \rangle_{\mathcal{H}}.$$

Now recalling Definition 2.32 we have $|v\rangle(1) = v$ and $|w\rangle(\alpha) = \alpha w$. Combined with the linearity of the inner product in its second argument we obtain

$$\langle |v\rangle(1), |w\rangle(\alpha) \rangle_{\mathcal{H}} = \alpha \langle v, w \rangle_{\mathcal{H}}$$

as required. \square

As a consequence of this proposition, in the following we use the notation $\langle v|w\rangle = \langle v,w\rangle$ (essentially “collapsing” the vertical lines of the bra and the ket to form a *bracket*). Also, we will denote elements of a Hilbert space by the associated ket, that is, we will usually write $|v\rangle \in \mathcal{H}$. Perhaps somewhat confusingly at first, we will actually use the ket notation either to mean an element of \mathcal{H} , or a map $\mathbb{C} \rightarrow \mathcal{H}$. While this may sound like a slight inconvenience, actually one of the strengths of the Dirac notation is that it uses this ambiguity to its advantage: the distinction between an element of \mathcal{H} and a map $\mathbb{C} \rightarrow \mathcal{H}$ is not important in the context we are concerned with, and this more relaxed notation makes certain algebraic manipulation easier. (Just think of the convenient notation for derivatives, $\frac{dy}{dx}$, and how physicists enjoy “multiplying by dx ” from time to time. Whenever one has a solid understanding of how this works and when this can be used, it’s a rather convenient thing to do.)

We can also compose a ket $|v\rangle : \mathbb{C} \rightarrow \mathcal{K}$ with a bra $\langle w| : \mathcal{H} \rightarrow \mathbb{C}$ to obtain an operator $|v\rangle\langle w| : \mathcal{H} \rightarrow \mathcal{K}$ (sometimes referred to as a *ket-bra*). According to the above, this operator is given by its action $|v\rangle\langle w|(|\psi\rangle) = |v\rangle\langle w|\psi\rangle = \langle w|\psi\rangle |v\rangle$ for all $|\psi\rangle \in \mathcal{H}$. The usual rules for composition of operators then apply, e.g. in the case of $\mathcal{H} = \mathcal{K}$, $(AB)^\dagger = B^\dagger A^\dagger$ gives $(|v\rangle\langle w|)^\dagger = |w\rangle\langle v|$.

We will mostly be concerned with operators on a single Hilbert space \mathcal{H} (as opposed to operators from \mathcal{H} to \mathcal{K}) and we will often make extensive use of the Dirac notation when describing operators. As a central example, consider an orthonormal basis $\{|e_i\rangle\}_{i=1}^d$ of a d -dimensional Hilbert space \mathbb{C}^d , and the operators $|e_i\rangle\langle e_j|$. Recalling that every operator on \mathbb{C}^d is given by a matrix $M_{d \times d}(\mathbb{C})$, we have that in the basis $\{|e_i\rangle\}$, the elements of the matrix associated to $|e_i\rangle\langle e_j|$ are given by

$$(|e_i\rangle\langle e_j|)_{k,l} = \langle e_k|(|e_i\rangle\langle e_j|)|e_l\rangle = \langle e_k|e_i\rangle\langle e_j|e_l\rangle = \delta_{i,k}\delta_{j,l},$$

where $\delta_{i,k}$ is the Kronecker delta. That is, the (i,j) element of the matrix is 1, and the rest are 0. This naturally implies that every operator A on \mathbb{C}^d can be written as

$$A = \sum_{i,j=1}^d a_{ij} |e_i\rangle\langle e_j|,$$

where $\{|e_i\rangle\}$ is an orthonormal basis of \mathcal{H} , and a_{ij} are complex numbers. Note that this is a manifestation of the fact that the set of operators on \mathbb{C}^d forms a vector space (in fact, a Hilbert space) of dimension d^2 , and a basis of this vector space is given by the operators $\{|e_i\rangle\langle e_j|\}_{i,j=1}^d$.

TERMINOLOGY & NOTATION. Usually we choose a fixed basis of \mathbb{C}^d that we call the *computational basis* (or *canonical basis*, or *standard basis*). The most common notation in quantum information theory is $\{|j\rangle\}_{j=0}^{d-1}$, noting that the indexing starts from 0 instead of 1, which is common in information theory. As an example, the computational basis of \mathbb{C}^2 is denoted by $\{|0\rangle, |1\rangle\}$. \diamond

EXAMPLE 2.36. Consider the Hilbert space \mathbb{C}^2 and the computational basis $\{|0\rangle, |1\rangle\}$. In the computational basis, the basis elements are given by

$$|0\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad |1\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

An arbitrary vector $|\psi\rangle \in \mathbb{C}^2$ in the computational basis is given by

$$|\psi\rangle = \alpha |0\rangle + \beta |1\rangle = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}$$

for $\alpha, \beta \in \mathbb{C}$. An example of a linear operator, $M \in \mathcal{L}(\mathbb{C}^2)$, in the computational basis is

$$M = |0\rangle\langle 1| = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix},$$

and an arbitrary linear operator, $N \in \mathcal{L}(\mathbb{C}^2)$, is written in the computational basis as

$$N = \alpha |0\rangle\langle 0| + \beta |0\rangle\langle 1| + \gamma |1\rangle\langle 0| + \delta |1\rangle\langle 1| = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}$$

for $\alpha, \beta, \gamma, \delta \in \mathbb{C}$. ◇

2.2.3 Spectral decomposition, positive semidefinite operators

Before stating the measurement postulate, we review some useful classes of operators—now using the Dirac notation.

TERMINOLOGY & NOTATION. In Dirac notation, for any operator $A \in \mathcal{L}(\mathcal{H}, \mathcal{K})$ and vectors $|\psi\rangle \in \mathcal{H}$ we have

$$A|\psi\rangle = |A\psi\rangle \in \mathcal{K}$$

and

$$\langle A\psi| = \langle \psi|A^\dagger \in \mathcal{L}(\mathcal{K}, \mathbb{C})$$
 ◇

DEFINITION 2.38. An operator $P \in \mathcal{L}(\mathcal{H})$ is an *orthogonal projection* if $P^2 = P = P^\dagger$. ◇

EXAMPLE 2.39. Consider the computational basis $\{|j\rangle\}_{j=0}^{d-1}$ on \mathbb{C}^d . Then

$$P_j = |j\rangle\langle j|$$

are orthogonal projections. We often say that “ P_j projects onto $|j\rangle$ ”, because $P_j|\psi\rangle$ is proportional to $|j\rangle$ for all $|\psi\rangle \in \mathbb{C}^d$. ◇

DEFINITION 2.40. The *rank* of an operator $A \in \mathcal{L}(\mathcal{H}, \mathcal{K})$ is the dimension of its *image*. That is, the dimension of the vector space spanned by the set

$$\text{Im}(A) := \{A|\psi\rangle : |\psi\rangle \in \mathcal{H}\} \subseteq \mathcal{K}.$$

Equivalently, the rank of A is the number of linearly independent columns of the matrix representation of A , or the number of linearly independent elements of the set $\{A|e_i\rangle : \{|e_i\rangle\}_i \text{ is a basis of } \mathcal{H}\}$. ◇

EXAMPLE 2.41. The rank of the projection $|j\rangle\langle j|$ is 1, and therefore such projections are usually called *rank-1 projections*. For any set of n orthonormal vectors, $\{|e_i\rangle\}_{i=1}^n \subset \mathcal{H}$, the operator

$$P = \sum_{i=1}^n |e_i\rangle\langle e_i|$$

is again a projection with rank n . \diamond

The following theorem provides a convenient way of writing Hermitian operators.

THEOREM 2.42 (SPECTRAL THEOREM). *Let $A \in \mathcal{L}(\mathcal{H})$ be a Hermitian (self-adjoint) operator on a d -dimensional Hilbert space \mathcal{H} . Then there exists an orthonormal basis $\{|e_i\rangle\}_{i=1}^d$ of \mathcal{H} consisting of eigenvectors of A . Furthermore, A can be written as*

$$A = \sum_{i=1}^d \lambda_i |e_i\rangle\langle e_i|,$$

where λ_i are the eigenvalues of A , that is, $A|e_i\rangle = \lambda_i|e_i\rangle$.

REMARK 2.43. Note that in the above description, λ_i can be zero. The number (including multiplicities) of non-zero eigenvalues of a Hermitian operator is also its rank, often denoted by $\text{rk}(A)$. Then A can also be written as

$$A = \sum_{i=1}^{\text{rk}(A)} \lambda_i |e_i\rangle\langle e_i|,$$

where the summation is over i such that $\lambda_i \neq 0$. \diamond

EXAMPLE 2.44. The identity operator on $\mathcal{H} = \mathbb{C}^d$ is Hermitian, with all of its d eigenvalues equal 1, and any vector $|\psi\rangle \in \mathcal{H}$ is an eigenvector. Therefore, for any orthonormal basis $\{|e_i\rangle\}_{i=1}^d$ the identity operator can be written as

$$\mathbb{1} = \sum_{i=1}^d |e_i\rangle\langle e_i|. \quad \diamond$$

Many objects in quantum (information) theory correspond to a special type of Hermitian operator called a positive semidefinite operator.

DEFINITION 2.45. An operator $A \in \mathcal{L}(\mathcal{H})$ is called *positive semidefinite*, denoted by $A \geq 0$ if

$$\langle \psi | A | \psi \rangle \geq 0 \quad \forall |\psi\rangle \in \mathcal{H}. \quad \diamond$$

REMARK 2.46. If the inequality is strict, that is, $\langle \psi | A | \psi \rangle > 0$ for all $|\psi\rangle \in \mathcal{H}$ then A is called *positive* or *positive definite*. \diamond

PROPOSITION 2.47. *If $A \in \mathcal{L}(\mathcal{H})$ is positive semidefinite then it is also Hermitian.*

PROOF. First we use the *polarisation identity*: for any $A \in \mathcal{L}(\mathcal{H})$ and all $|\psi\rangle, |\varphi\rangle \in \mathcal{H}$ we have

$$\begin{aligned}\langle A\psi|\varphi\rangle &= \frac{1}{4}[\langle A(\psi + \varphi)|\psi + \varphi\rangle - \langle A(\psi - \varphi)|\psi - \varphi\rangle \\ &\quad - i\langle A(\psi + i\varphi)|\psi + i\varphi\rangle + i\langle A(\psi - i\varphi)|\psi - i\varphi\rangle].\end{aligned}$$

Furthermore, for all $|x\rangle \in \mathcal{H}$ and a positive semidefinite operator A , we have

$$\langle Ax|x\rangle = \overline{\langle x|Ax\rangle} = \overline{\langle x|A|x\rangle} = \langle x|A|x\rangle = \langle x|Ax\rangle$$

and therefore $\langle A\psi|\varphi\rangle = \langle \psi|A\varphi\rangle$ for all $|\psi\rangle, |\varphi\rangle \in \mathcal{H}$ and thus $A^\dagger = A$. \square

DEFINITION 2.48. For two operators, $A, B \in \mathcal{L}(\mathcal{H})$, we say that A is *greater than* B , denoted $A \geq B$ if $A - B \geq 0$. This is in fact a *partial order* on $\mathcal{L}(\mathcal{H})$, that is,

1. $A \geq A$ for all $A \in \mathcal{L}(\mathcal{H})$ (reflexivity)
2. if $A \geq B$ and $B \geq A$ then $A = B$ (anti-symmetry)
3. if $A \geq B$ and $B \geq C$ then $A \geq C$ (transitivity)

Note that this is not a *total order*, that is, there exist $A, B \in \mathcal{L}(\mathcal{H})$ such that neither $A \geq B$ nor $B \geq A$ (even if they are both Hermitian). \diamond

EXAMPLE 2.49. Every orthogonal projection is positive semidefinite, since for $P^2 = P = P^\dagger$ we have

$$\langle \psi|P|\psi\rangle = \langle \psi|P^2|\psi\rangle = \langle P\psi|P\psi\rangle = \|P|\psi\rangle\|^2 \geq 0 \quad \forall |\psi\rangle \in \mathcal{H}. \quad \diamond$$

EXAMPLE 2.50. For any operator $A \in \mathcal{L}(\mathcal{H}, \mathcal{K})$, the operator $A^\dagger A$ is positive semidefinite, since

$$\langle \psi|A^\dagger A|\psi\rangle = \langle A\psi|A\psi\rangle = \|A|\psi\rangle\|^2 \geq 0.$$

In fact, an operator $M \in \mathcal{L}(\mathcal{H})$ is positive semidefinite if and only if it can be written as

$$M = A^\dagger A$$

for some $A \in \mathcal{L}(\mathcal{H}, \mathcal{K})$. \diamond

2.2.4 Measurement postulate

We are now ready to state the measurement postulate.

POSTULATE (MEASUREMENTS). Consider a physical system with an associated Hilbert space \mathcal{H} . A measurement with n possible outcomes corresponds to a collection of n operators (one representing each outcome) on \mathcal{H} , that is, a set $\{M_j\}_{j=1}^n$ with $M_j \in \mathcal{L}(\mathcal{H})$.

If the state of the system before the measurement is $|\psi\rangle \in \mathcal{H}$, then the probability of obtaining an outcome j is given by

$$p(j) = \langle \psi | M_j^\dagger M_j | \psi \rangle.$$

After observing outcome j , the updated state of the system is given by

$$|\varphi_j\rangle = \frac{M_j |\psi\rangle}{\sqrt{\langle \psi | M_j^\dagger M_j | \psi \rangle}} = \frac{M_j |\psi\rangle}{\sqrt{p(j)}}.$$

The measurement operators must satisfy the *completeness relation*,

$$\sum_{j=1}^n M_j^\dagger M_j = \mathbb{1},$$

reflecting the fact that the outcome probabilities form a proper probability distribution,

$$\sum_{j=1}^n p(j) = \sum_{j=1}^n \langle \psi | M_j^\dagger M_j | \psi \rangle = \langle \psi | \psi \rangle = 1$$

while we already have that $p(j) \geq 0$, since $M_j^\dagger M_j \geq 0$. \diamond

Therefore, a measurement—as required—takes as an input a quantum state $|\psi\rangle \in \mathcal{H}$ and as an output, gives us the probabilities $p(j)$ with which the different outcomes occur, as well as the states $|\varphi_j\rangle$ after observing outcome j .

In quantum information theory—and in much of this module—we only care about the outcome probabilities and not necessarily the state of the system after the measurement. This gives rise to a more abstract notion of a quantum measurement:

DEFINITION 2.51. If we only care about the outcome probabilities, a quantum measurement with n outcomes is described by a *positive-operator-valued measure* (POVM). A POVM corresponds to a set of n positive semidefinite operators adding up to the identity operator, that is, $\{N_j\}_{j=1}^n \subset \mathcal{L}(\mathcal{H})$ such that

$$N_j \geq 0 \quad \text{and} \quad \sum_j N_j = \mathbb{1}.$$

If the state before the measurement is described by $|\psi\rangle \in \mathcal{H}$, then the outcome probabilities are given by $p(j) = \langle \psi | N_j | \psi \rangle$, which indeed forms a valid probability distribution. The operators N_j are often called *POVM elements* or sometimes *measurement operators* or *effects*. \diamond

REMARK 2.52. The POVM corresponding to the measurement from Postulate 6 is given by $N_j = M_j^\dagger M_j$. \diamond

REMARK 2.53. In quantum *mechanics* (and sometimes also in quantum information), the measurement outcome is usually a physical quantity—a real number (for example, the position x of a particle). Therefore, in quantum

mechanics, measurements are usually associated to an *observable* $A \in \mathcal{L}(\mathcal{H})$, which is a Hermitian operator, and the eigenvalues of the operator are the possible outcomes. Assuming that \mathcal{H} is d -dimensional, due to Theorem 2.42, this operator can be written as

$$A = \sum_{i=1}^d \lambda_i |e_i\rangle\langle e_i|$$

for some $\lambda_i \in \mathbb{R}$ and an orthonormal basis $\{|e_i\rangle\}_{i=1}^d$ of \mathcal{H} . If the state of the system is described by $|\psi\rangle \in \mathcal{H}$, then the probability of observing outcome λ_i is given by $|\langle\psi|e_i\rangle|^2 = \langle\psi|e_i\rangle\langle e_i|\psi\rangle$. If we only care about the probabilities, then A can be associated to a d -outcome measurement with POVM elements $|e_i\rangle\langle e_i|$. \diamond

An important class of POVMs is *projective measurements*.

DEFINITION 2.54. A POVM $\{P_j\}_{j=1}^n$ is called a *projective measurement* or *projection-valued measure* (PVM) if $P_j^2 = P_j$ for all $j \in \{1, \dots, n\}$. \diamond

EXAMPLE 2.55. Consider an orthonormal basis $\{|e_j\rangle\}_{j=1}^d$ of a d -dimensional Hilbert space. Then the set

$$\{P_j = |e_j\rangle\langle e_j|\}_{j=1}^d$$

defines a POVM. In fact, all the operators are rank-1 projections, and therefore it is often called a *rank-1 projective measurement*, but also a *von Neumann measurement*, or a *measurement in the basis* $\{|e_j\rangle\}_{j=1}^d$. \diamond

EXAMPLE 2.56. Consider the measurement defined by the set $\{M_1 = \mathbb{1}, M_2 = 0, M_3 = 0\}$. This is a three-outcome measurement, but for any state the outcome is always “1”, that is, $p(j) = \delta_{j,1}$ for all j and all states $|\psi\rangle \in \mathcal{H}$. \diamond

EXAMPLE 2.57. Consider the n -outcome measurement defined by the set $\{\frac{1}{n}, \frac{1}{n}, \dots, \frac{1}{n}\}$. This measurement provides a completely random output independently of the state, that is, $p(j) = \frac{1}{n}$ for all j and all states $|\psi\rangle \in \mathcal{H}$. \diamond

EXAMPLE 2.58. Consider the rank-1 projective measurement on \mathbb{C}^2 given by $\{|0\rangle\langle 0|, |1\rangle\langle 1|\}$ (the measurement in the computational basis). If the measured system is given by $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$, then the outcome probabilities are given by

$$\begin{aligned} p(0) &= (\bar{\alpha}\langle 0| + \bar{\beta}\langle 1|) |0\rangle\langle 0| (\alpha|0\rangle + \beta|1\rangle) = |\alpha|^2 \\ p(1) &= (\bar{\alpha}\langle 0| + \bar{\beta}\langle 1|) |1\rangle\langle 1| (\alpha|0\rangle + \beta|1\rangle) = |\beta|^2 \end{aligned}$$

The post-measurement states are given by

$$\begin{aligned} |\varphi_0\rangle &= \frac{|0\rangle\langle 0|(\alpha|0\rangle + \beta|1\rangle)}{\sqrt{p(0)}} = \frac{\alpha}{|\alpha|} |0\rangle = e^{i\theta_\alpha} |0\rangle \\ |\varphi_1\rangle &= \frac{|1\rangle\langle 1|(\alpha|0\rangle + \beta|1\rangle)}{\sqrt{p(1)}} = \frac{\beta}{|\beta|} |1\rangle = e^{i\theta_\beta} |1\rangle \end{aligned}$$

where $\alpha = |\alpha| e^{i\theta_\alpha}$ and $\beta = |\beta| e^{i\theta_\beta}$. That is, after observing the outcome “0”, the state is equivalent (up to a phase, i.e. a complex number with modulus 1) to $|0\rangle$, and after observing “1”, the state is equivalent to $|1\rangle$. If no other evolution takes place, after first observing outcome “0”, repeated measurements will always yield outcome “0”. This is a generic property of projective measurements. \diamond

2.3 DYNAMICS

We now know how to mathematically represent the state of a quantum system, and how to extract some information from the system. But much of physics (and also quantum information, as we will see) is about *dynamics*, that is, the time evolution of physical systems. If I know that my state is given by $|\psi\rangle$ at time t , what will be the state at some later time t' ? This question can be important e.g. in quantum computation, as quantum computation involves three main steps: *i*) *encoding* some classical information (e.g. a bit string, an input to a function) into a quantum state; *ii*) *evolving* the state (this is the main part of the computation, done on the quantum level); *iii*) *measuring* the evolved state (turning the quantum information back to classical information, e.g. a bit string, the output of the function).

In quantum theory—given our knowledge so far—this evolution should be described by a map $U : \mathcal{H} \rightarrow \mathcal{H}$: it maps the state at time t to the state at time t' . This map should also be *linear* (i.e. a linear operator), which can be traced back to the wave-mechanical origins of quantum theory. Thus, we are looking for a linear operator, $U \in \mathcal{L}(\mathcal{H})$ that maps every quantum state to a quantum state, that is, $\|U|\psi\rangle\| = \||\psi\rangle\|$ for all $|\psi\rangle \in \mathcal{H}$ (a *length-preserving operator*, in particular mapping normalised vectors to normalised vectors). These operators turn out to be equivalent to *unitary* operators.

DEFINITION 2.59. An operator $U \in \mathcal{L}(\mathcal{H})$ is called *unitary* if

$$U^\dagger U = \mathbb{1}, \quad \text{or, equivalently,} \quad U U^\dagger = \mathbb{1}. \quad \diamond$$

To see the equivalence above, it is worth taking a look at the concept of an *inverse*.

DEFINITION 2.60. Consider an operator $A \in \mathcal{L}(\mathcal{H})$. If there exists an operator $A_L \in \mathcal{L}(\mathcal{H})$ such that

$$A_L^{-1} A = \mathbb{1},$$

then A_L^{-1} is called the *left inverse* of A . Similarly, if there exists an operator $A_R^{-1} \in \mathcal{L}(\mathcal{H})$ such that

$$A A_R^{-1} = \mathbb{1},$$

then A_R^{-1} is called the *right inverse* of A . If there exists an operator $A^{-1} \in \mathcal{L}(\mathcal{H})$ such that

$$A^{-1} A = A A^{-1} = \mathbb{1},$$

i.e. A^{-1} is both the right and left inverse of A , then A^{-1} is called the *inverse* of A , and A is called *invertible*. \diamond

THEOREM 2.61. *In finite dimensions, if there exists a left inverse, it is also a right inverse, and therefore it is the inverse. Moreover, this inverse is unique.*

PROPOSITION 2.62. *A unitary operator is invertible.*

PROOF. By definition, U^\dagger is both a left and right inverse of U and thus an inverse. \square

We now prove that unitary operators are the same as length-preserving operators.

PROPOSITION 2.63. *An operator $U \in \mathcal{L}(\mathcal{H})$ is unitary if and only if it is length-preserving⁴ (i.e., maps states to states) and invertible.⁵*

⁴ Such operators are also called isometric.

PROOF. For the “only if” direction, take an arbitrary state $|\psi\rangle \in \mathcal{H}$ and assume that $U \in \mathcal{L}(\mathcal{H})$ is unitary. Then we have

⁵ Actually, in finite dimensions, length-preserving implies invertible.

$$\|\psi\|^2 = \langle \psi | \psi \rangle = \langle \psi | U^\dagger U \psi \rangle = \langle U \psi | U \psi \rangle = \|U |\psi\rangle\|^2,$$

that is, U is length-preserving.

For the “if” direction, assume that $U \in \mathcal{L}(\mathcal{H})$ is length-preserving. Then for all $|\psi\rangle \in \mathcal{H}$ we have

$$0 = \langle U \psi | U \psi \rangle - \langle \psi | \psi \rangle = \langle \psi | (U^\dagger U - \mathbb{1}) \psi \rangle,$$

and thus $U^\dagger U = \mathbb{1}$ (remember the polarisation identity!). Since U is invertible and U^\dagger is a left inverse, it must be the inverse and therefore also $U U^\dagger = \mathbb{1}$ is satisfied. \square

We can now state the postulate concerning the dynamics of quantum systems.

POSTULATE (DYNAMICS). The time evolution of a *closed* quantum system associated to a Hilbert space \mathcal{H} is described by a unitary operator $U \in \mathcal{L}(\mathcal{H})$. That is, the state $|\psi(t)\rangle$ of a system at time t is related to the state $|\psi(t')\rangle$ of the system at time t' by a unitary operator U that only depends on t and t' ,

$$|\psi(t')\rangle = U |\psi(t)\rangle. \quad \diamond$$

REMARK 2.64. The postulate only holds if the system is *closed*, that is, it does not interact with any other systems. Of course, in reality this is essentially impossible, but the theory describes an idealised scenario. Moreover, every system (even if not closed) can be considered as part of a bigger closed system (e.g. the whole Universe) that evolves unitarily. The evolution of our system can then be recovered from the overall unitary evolution. \diamond

Unitary operators play an important role in quantum computing: *quantum gates*, the operations that replace the logic gates from classical circuits, are represented by unitary operators, since they represent the time evolution of the quantum state during the computation.

An important set of operators are the *Pauli matrices*

$$X := |0\rangle\langle 1| + |1\rangle\langle 0|, \quad (2.4a)$$

$$Y := -i|0\rangle\langle 1| + i|1\rangle\langle 0|, \quad (2.4b)$$

$$Z := |0\rangle\langle 0| - |1\rangle\langle 1|. \quad (2.4c)$$

In the computational basis (with standard ordering) they can be represented by the matrices

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

The X gate is similar to a classical NOT gate and ‘flips’ a $|0\rangle$ to $|1\rangle$ and vice versa:

$$X|0\rangle = |1\rangle, \quad X|1\rangle = |0\rangle.$$

The Y and Z gate have no classical analogue. Their action on the vectors of the computational basis is:

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

Before mentioning other widely used unitary operators, it is useful to introduce the concept of *superposition*.

DEFINITION 2.65. Let us fix an orthonormal basis $\{|e_i\rangle\}_{i=1}^d$ of a d -dimensional Hilbert space \mathcal{H} . We say that a vector (or usually a state) $|\psi\rangle \in \mathcal{H}$ is a *superposition* of these basis elements if in the unique decomposition

$$|\psi\rangle = \sum_{i=1}^d \alpha_i |e_i\rangle$$

at least two $\alpha_i \in \mathbb{C}$ are non-zero. \diamond

REMARK 2.66. Note that the concept of superposition *depends on the basis!* For example, the (often used) state

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

is in a superposition in the basis $\{|0\rangle, |1\rangle\}$. However, if we write it in the Fourier (or Hadamard) basis, $\{|+\rangle, |-\rangle\} = \{\frac{1}{\sqrt{2}}(|0\rangle + |1\rangle), \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle)\}$ then it is no longer in a superposition. \diamond

REMARK 2.67. Superposition does not have a classical physical analogue. A superposition of the form $\sqrt{p}|0\rangle + \sqrt{1-p}|1\rangle$ does represent a system that is “in the state $|0\rangle$ with probability p ” and “in the state $|1\rangle$ with probability $1-p$ ” in the sense that if we measure this state in the $\{|0\rangle\langle 0|, |1\rangle\langle 1|\}$ basis, we

obtain outcome “0” with probability p and outcome “1” with probability $1 - p$. However, this state does *not* represent the physical scenario of “preparing state $|0\rangle$ with probability p and state $|1\rangle$ probability $1 - p$ ” (which would be a “classical” mixture). Therefore, the uncertainty in the measurement outcome does *not* come from our limited knowledge (maybe someone else prepared the states with probabilities p and $1 - p$ and we just don’t know which one we have), but it is *intrinsic* to quantum theory. We will see in later sections how to represent classical mixtures in a quantum mechanical framework. \diamond

Another important unitary operator is the *Hadamard operator*. It is defined as

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

In the computational basis (with standard ordering) it is represented by the matrix

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

The Hadamard gate is useful to produce superpositions of $|0\rangle$ and $|1\rangle$:

$$H|0\rangle = |+\rangle, \quad H|1\rangle = |-\rangle.$$

REMARK 2.68. Note that quantum measurements also represent a kind of time evolution: the state of a quantum system changes after performing a measurement on it (e.g. from $|\psi\rangle$ to $M_m |\psi\rangle / \sqrt{\langle \psi | M_m^\dagger M_m | \psi \rangle}$ if we obtain outcome m of measurement M on a quantum state $|\psi\rangle$). This evolution is, however, *not* unitary in general. This tension between the usual unitary evolution and the evolution induced by a measurement is what is often referred to as *the measurement problem* of quantum theory, and it is one of the central problems in the foundations and interpretations of quantum theory. This problem is beyond the scope of this module, and by “time evolution”, we will usually understand a unitary transformation. \diamond

3 Qubits

Qubits (quantum bits) are the simplest non-trivial quantum systems, associated to the Hilbert space \mathbb{C}^2 . As we have seen before, an arbitrary qubit state can be written in the computational basis as

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

where $|\alpha|^2 + |\beta|^2 = 1$. Qubits are the quantum analogues of *bits*, the basic units of *classical* information processing. A classical bit can only be in two states: 0 or 1. On the other hand, a qubit can be in any superposition of the states $|0\rangle$ and $|1\rangle$, giving rise to a much richer structure than that of bits. This rich structure—together with other quantum phenomena that we will discuss later in the module—is one of the features that allows quantum information processing protocols (such as computation or cryptography where information processing is done using qubits) to outperform their classical counterparts.

3.1 BLOCH SPHERE

The set of qubit states have a convenient geometrical representation called the *Bloch sphere*. Since this is a useful tool both for calculations and for visualisation, in this section we review the notion of the Bloch sphere.

First, remember that two states, $|\psi\rangle, |\varphi\rangle \in \mathbb{C}^2$ are *equivalent* if

$$|\varphi\rangle = e^{i\theta} |\psi\rangle$$

for some $\theta \in \mathbb{R}$. The proportionality factor $e^{i\theta}$ is called a *global phase*.

While the global phase is not relevant in quantum theory, the *relative phase* of a qubit state is physically significant. The relative phase of a superposition $\alpha |v\rangle + \beta |w\rangle$ of two orthogonal vectors $|v\rangle, |w\rangle$ is the complex number $e^{i\delta}$ given by

$$\frac{\alpha}{\beta} = \frac{e^{i\phi} |\alpha|}{e^{i\theta} |\beta|} = e^{i\delta} \frac{|\alpha|}{|\beta|}.$$

Note that two states $\alpha |v\rangle + \beta |w\rangle$ and $\alpha' |v\rangle + \beta' |w\rangle$ where the amplitudes have the same modulus ($|\alpha| = |\alpha'|$ and $|\beta| = |\beta'|$) but differ in relative phase represent different states. As an example, the following qubit states will appear repeatedly in this module:

$$\begin{aligned} |+\rangle &:= \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle), & |+\mathrm{i}\rangle &:= \frac{1}{\sqrt{2}}(|0\rangle + \mathrm{i}|1\rangle), \\ |-\rangle &:= \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle), & |-\mathrm{i}\rangle &:= \frac{1}{\sqrt{2}}(|0\rangle - \mathrm{i}|1\rangle). \end{aligned}$$

Note that all of these vectors differ only by their relative phase, but some of them are actually orthogonal, as both $\{|+\rangle, |-\rangle\}$ and $\{|+i\rangle, |-i\rangle\}$ form orthonormal bases of \mathbb{C}^2 . Therefore, these states are not equivalent to each other (they don't differ by a global phase, but by their relative phases).

While the relative phase is physically significant, the global phase is not (by definition). Therefore, when we characterise a quantum state, it's sufficient to characterise it up to a global phase. The often used convention is to use this freedom in the global phase to set the coefficient of the $|0\rangle$ basis element to a non-negative real number. Let us consider an arbitrary qubit state,

$$|v\rangle = |\alpha| e^{i\varphi_\alpha} |0\rangle + |\beta| e^{i\varphi_\beta} |1\rangle.$$

where instead of $\alpha, \beta \in \mathbb{C}$, we wrote these complex numbers in their polar form, $\alpha = |\alpha| e^{i\varphi_\alpha}$ and $\beta = |\beta| e^{i\varphi_\beta}$. This state is equivalent to

$$|\tilde{v}\rangle = e^{-i\varphi_\alpha} |v\rangle = |\alpha| |0\rangle + |\beta| e^{i(\varphi_\beta - \varphi_\alpha)} |1\rangle, \quad (3.1)$$

where now the coefficient of $|0\rangle$ is a non-negative real number, $|\alpha|$. The normalisation condition translates to

$$1 = \langle \tilde{v} | \tilde{v} \rangle = |\alpha|^2 + |\beta|^2.$$

Therefore, we can parametrise $|\alpha|$ and $|\beta|$ by a single angle: by convention, we denote this angle by $\frac{\theta}{2}$, and we can write $\frac{\theta}{2} = \arccos(|\alpha|)$, or equivalently, $|\alpha| = \cos \frac{\theta}{2}$, which also implies that $|\beta| = \sin \frac{\theta}{2}$, since $1 = |\alpha|^2 + |\beta|^2 = \cos^2 \left(\frac{\theta}{2}\right) + \sin^2 \left(\frac{\theta}{2}\right)$. Note that since $|\alpha| \in [0, 1]$, we have $\frac{\theta}{2} \in [0, \frac{\pi}{2}]$. We further denote the angle $\varphi_\beta - \varphi_\alpha$ in Eq. (3.1) by ϕ . Then we have that every state $|v\rangle \in \mathbb{C}^2$ is equivalent to a state of the form

$$|\tilde{v}\rangle = \cos \frac{\theta}{2} |0\rangle + \sin \frac{\theta}{2} e^{i\phi} |1\rangle.$$

That is, every state (up to a global phase) can be represented by an ordered pair of two angles, (θ, ϕ) , where $\theta \in [0, \pi]$ and $\phi \in [0, 2\pi)$.

PROPOSITION 3.1. *Every qubit state is equivalent to*

$$|v\rangle = \cos \frac{\theta}{2} |0\rangle + e^{i\phi} \sin \frac{\theta}{2} |1\rangle \quad (3.2)$$

for some $\theta \in [0, \pi]$ and $\phi \in [0, 2\pi)$. Except when $|v\rangle = |0\rangle$ or $|v\rangle = |1\rangle$, this representation is unique.

REMARK 3.2. When $|v\rangle = |0\rangle$, every pair $(0, \phi)$ represents $|v\rangle$, and when $|v\rangle = |1\rangle$, every pair (π, ϕ) represents $|v\rangle$. Otherwise, the correspondence between a qubit state $|v\rangle$ (up to a global phase) and a pair (θ, ϕ) is one-to-one. \diamond

Note that any pair (θ, ϕ) with $\theta \in [0, \pi]$ and $\phi \in [0, 2\pi)$ represents a unique point on the unit sphere embedded in three dimensions (again, this point is unique unless $\theta = 0$ or $\theta = \pi$). As such, the representation of $|v\rangle$ from Proposition 3.1 is called the *Bloch sphere representation* of $|v\rangle$ (see Figure 3.1).

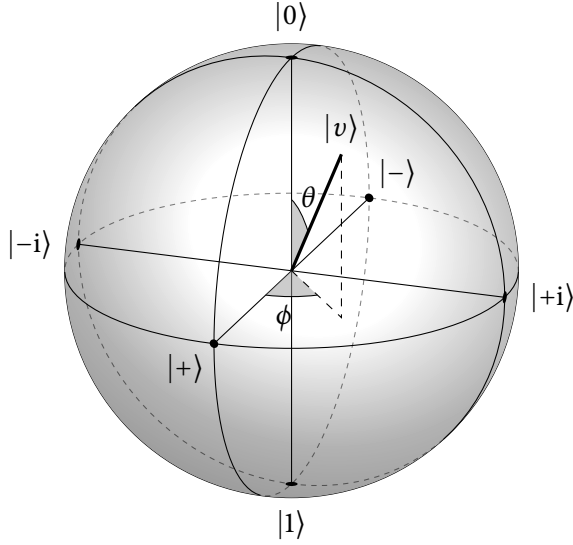


FIGURE 3.1. An illustration of the Bloch sphere. A point (θ, ϕ) on the sphere corresponds to the state $|v\rangle = \cos \frac{\theta}{2} |0\rangle + e^{i\phi} \sin \frac{\theta}{2} |1\rangle$.

The collection of vectors of the form in Eq. (3.2) form the Bloch sphere. For example, the state vectors introduced earlier are mapped onto the Bloch sphere (parametrised by (θ, ϕ)) as

$$\begin{aligned} |0\rangle &\mapsto (0, 0), & |+ \rangle &\mapsto (\frac{\pi}{2}, 0), & |+i \rangle &\mapsto (\frac{\pi}{2}, \frac{\pi}{2}), \\ |1\rangle &\mapsto (\pi, 0), & |- \rangle &\mapsto (\frac{\pi}{2}, \pi), & |-i \rangle &\mapsto (\frac{\pi}{2}, \frac{3\pi}{2}). \end{aligned}$$

Using the Bloch sphere it is easy to read off all possible states, and their relations have simple geometric interpretations. As an example, a rank-1 projective measurement on a qubit corresponds to measuring in some orthonormal basis $\{|\varphi\rangle, |\varphi^\perp\rangle\}$. As such, describing orthogonal vectors on the Bloch sphere is directly relevant when describing measurements.

PROPOSITION 3.3. *Two qubit states are orthogonal if and only if they are represented by antipodal points on the Bloch sphere.*

Therefore, a rank-1 projective measurement is described by a single point (θ, ϕ) on the Bloch sphere. The two measurement operators are the projections onto the state represented by (θ, ϕ) and its antipodal point, $(\pi - \theta, \phi + \pi)$.

The evolution of qubit states can also be described on the Bloch sphere. In later sections we will see that a unitary operation corresponds to a rotation on the Bloch sphere.

3.2 BB84 PROTOCOL

In this section, we take a look at an actual quantum information processing protocol, using what we have learnt about quantum states and measurements, specifically in the qubit case. The protocol is a cryptographic protocol known as *key distribution*.

One of the difficulties in establishing a secure communication channel using encryption is that of sharing the decryption key. There exist two different types of cryptographic schemes: *symmetric* and *asymmetric cryptography*.

In symmetric cryptography, only one *shared key* exists and is used for both encryption and decryption. Before the secure communication can be initiated, the shared key must be exchanged securely—a process which is often difficult to realise and prone to interception.

In asymmetric cryptography, there are two keys, called the *private* and the *public key*. The public key can be used by anyone to encrypt messages, but only the owner of the private key can decrypt them. Underlying an asymmetric cryptographic protocol is an asymmetric mathematical problem like that of multiplying large prime numbers and factorising such products—no fast (classical) method is known for the latter, while multiplication is trivial. The way this example is used in cryptography is as follows (roughly): consider two large prime numbers, p and q , and their product pq . The product is used as a public key, and can be used to encrypt messages. For the decryption, however, the knowledge of p and q is required. Cryptographic protocols of this type rely on the assumption that no fast algorithm exists that solves the underlying mathematical problem. However, these assumptions can turn out to be wrong: quantum computers implementing Shor's algorithm⁶ can be used to break cryptographic protocols that rely on prime number factorisation.

Quantum key distribution (QKD) is a symmetric cryptographic scheme that overcomes the problem of cryptographic protocols relying on computational assumptions (e.g. the assumption that factoring primes is difficult). The security of QKD is instead guaranteed by *the laws of quantum theory*, that is, by the way quantum states and measurements behave.

The aim of QKD is for two distant parties, Alice and Bob, to establish a shared cryptographic key. A shared key is a string of bits that is known to both Alice and Bob, but unknown to any potential eavesdropper. In 1984, Charles H. Bennett and Gilles Brassard proposed the first QKD protocol,⁷ usually referred to as the BB84 protocol. In this protocol, Alice and Bob try to establish a shared secret key using communication over a public, but authenticated classical channel and a single qubit quantum channel. An authenticated channel is a channel such that Alice and Bob can be certain that they are talking to each other (there is no one impersonating them), but eavesdroppers might be able to listen in. The eavesdropper (Eve) also has access to the quantum channel, that is, she can intercept, measure, and re-send quantum states sent through the channel. The protocol consists of the following steps:

BB84 QKD PROTOCOL.

1. Alice secretly chooses a string, a , of N random (classical) data bits (these will eventually be converted into the shared key). For example, $a = 01100010\dots$

⁶ P. W. Shor, “Algorithms for quantum computation: Discrete logarithms and factoring”. *Proceedings of the 35th Annual Symposium on Foundations of Computer Science. IEEE Comput. Soc. Press.* pp. 124–134, 1994.

⁷ C. H. Bennett, G. Brassard, “Quantum cryptography: Public key distribution and coin tossing”. *Proceedings of IEEE International Conference on Computers, Systems and Signal Processing*, pp. 175–179, 1984.

2. Additionally, Alice and Bob each secretly choose a random N -bit string a' and b' , respectively.
3. Alice encodes each data bit of a into a qubit in one of two possible ways: either as $\{|0\rangle, |1\rangle\}$ if the corresponding bit of a' is 0 (that is, she encodes $0 \mapsto |0\rangle$ and $1 \mapsto |1\rangle$), or as $\{|+\rangle, |-\rangle\}$ if a' is 1 (that is, she encodes $0 \mapsto |+\rangle$ and $1 \mapsto |-\rangle$).
4. Alice sends all of the resulting qubits to Bob.
5. Bob measures each of the received qubits in one of two possible ways: either in the basis $\{|0\rangle, |1\rangle\}$ if the corresponding bit of b' is 0, or in the basis $\{|+\rangle, |-\rangle\}$ if b' is 1. He saves the result in b .
6. Bob announces that he has received Alice's transmission.
7. Alice announces a' .
8. Bob announces which of the bits of a' and b' are not equal.
9. Alice and Bob discard any bits of a and b where a' and b' do not match.
10. Alice randomly selects a subset of the remaining bits and discloses her choice of this subset to Bob.
11. Alice and Bob announce the corresponding bits of a and b in this subset. If more than an acceptable number disagree, they abort the protocol. If they don't abort, they discard the disclosed bits (as they are not secure, being publicly announced).
12. Alice and Bob use the remaining bits to create a shared secret key.

A simulation of the BB84 protocol is presented in Table 3.1.

Alice's random bits a	0	1	1	0	1	1	0	0
Alice's random bases a'	1	0	1	0	0	0	0	1
Bob's random bases b'	0	1	1	1	1	0	0	1
Qubits Alice sends	$ +\rangle$	$ 1\rangle$	$ -\rangle$	$ 0\rangle$	$ 1\rangle$	$ 1\rangle$	$ 0\rangle$	$ +\rangle$
Bits b Bob measures	1	0	1	1	1	1	0	0
Comparison of bases	\neq	\neq	$=$	\neq	\neq	$=$	$=$	$=$
Remaining bits of a			1			1	0	0
Remaining bits of b			1			1	0	0
Comparison of selected bits						$=$		$=$
Shared secret bits			1				0	

TABLE 3.1. A 'simulation' of the BB84 protocol with $N = 8$. Note again that a measurement result from step 5. is deterministic if Alice's and Bob's corresponding bases agree and random if the bases disagree. After the comparison of basis vectors in step 9. there are 4 bits left. After the comparison of 2 bits in step 11, the 3rd and 7th bit are left to establish the secret key.

The fundamental idea behind this protocol is that the four states

$$\{|0\rangle, |1\rangle, |+\rangle, |-\rangle\}$$

are not all mutually orthogonal. It follows that if Eve intercepts Alice's qubit—which is in one of the four states above—she cannot infer with certainty

which state was sent. Indeed, to perfectly distinguish $|0\rangle$ from $|1\rangle$, Eve should measure in the $\{|0\rangle, |1\rangle\}$ basis, as her outcome will always be “0” if the state sent was $|0\rangle$ and always “1” if the state sent was $|1\rangle$. However, if the state sent was $|+\rangle$ or $|-\rangle$, Eve’s outcome will be “0” or “1” with probability $\frac{1}{2}$ each, revealing no information to Eve.

Furthermore, ideally Eve would like to go unnoticed, since if Alice and Bob detect her presence, they can abort their protocol, and there’s no key to intercept. In the above scenario, Eve can go unnoticed whenever the state was $|0\rangle$ or $|1\rangle$: she can simply re-send the state corresponding to her outcome (or equivalently, forward the post-measurement state). However, if the state was $|+\rangle$ or $|-\rangle$, Eve doesn’t know what state to send to Bob, and she will inevitably introduce errors in the data of Alice and Bob (the bit-strings a and b will not match in many cases). Too much interference from Eve’s side therefore leads to aborting the protocol.

A completely analogous argument holds when Eve chooses to measure in the $\{|+\rangle, |-\rangle\}$ basis, and it can be shown that choosing other measurements does not help her either. Ultimately, Eve’s problem is that she does not know what measurement to use in order to obtain information about the bit encoded by Alice.

A simulation of the same implementation of the BB84 protocol as above, but with Eve present and measuring in each round randomly either in the $\{|0\rangle, |1\rangle\}$ basis ($e' = 0$) or in the $\{|+\rangle, |-\rangle\}$ basis ($e' = 1$) is presented in Table 3.2.

Alice’s random bits a	0	1	1	0	1	1	0	0
Alice’s random bases a'	1	0	1	0	0	0	0	1
Bob’s random bases b'	0	1	1	1	1	0	0	1
Eve’s random bases e'	1	1	1	1	0	1	0	0
Qubits Alice sends	$ +\rangle$	$ 1\rangle$	$ -\rangle$	$ 0\rangle$	$ 1\rangle$	$ 1\rangle$	$ 0\rangle$	$ +\rangle$
Bits Eve measures e	0	1	1	1	1	1	0	0
Qubits Eve sends	$ +\rangle$	$ -\rangle$	$ -\rangle$	$ -\rangle$	$ 1\rangle$	$ -\rangle$	$ 0\rangle$	$ 0\rangle$
Bits b Bob measures	1	1	1	1	1	0	0	0
Comparison of bases	\neq	\neq	$=$	\neq	\neq	$=$	$=$	$=$
Remaining bits of a			1			1	0	0
Remaining bits of b			1			0	0	0
Comparison of selected bits						\neq		$=$

TABLE 3.2. A ‘simulation’ of the BB84 protocol with $N = 8$, with an eavesdropper. Note the differences in Bob’s measurements of the 2nd and 6th bit compared to Table 3.1. In the final comparison of step 11. Alice and Bob notice that they disagree about the 6th bit—this hints at an eavesdropper (or a noisy quantum channel).

4 Composite systems

In previous sections, we were discussing the mathematical description of physical systems. We saw that for every physical system, there exists an associated Hilbert space. But how do we describe *multiple* systems at the same time? This becomes essential if multiple physical systems (e.g. many electrons or photons) interact with one another. Furthermore, this is exactly what happens in quantum computing, where we would ideally like to control multiple qubits and let them interact so that we can perform efficient quantum computation. The system of multiple subsystems is called a *composite system*. We will see that composite systems in quantum theory have radically different properties than composite systems in classical physics. These differences are at the heart of quantum advantage in information processing tasks, and are also closely related to the foundations of quantum theory.

4.1 TENSOR PRODUCT

The next (and modulo some extensions that we will discuss in the next chapter, the last) postulate of quantum theory prescribes the Hilbert space associated to composite systems.

POSTULATE (COMPOSITE SYSTEMS). Consider a physical system associated to the Hilbert space \mathcal{H} , and another physical system associated to the Hilbert space \mathcal{K} . Then, the joint system is described by the *tensor product* of the two Hilbert spaces, $\mathcal{H} \otimes \mathcal{K}$. \diamond

In order to understand this postulate, we need to define the tensor product of Hilbert spaces.

DEFINITION 4.1. The *tensor product* $\mathcal{V} \otimes \mathcal{W}$ of two Hilbert spaces \mathcal{V} and \mathcal{W} is the Hilbert space spanned by elements of the form $|v\rangle \otimes |w\rangle$ subject to the relations⁸

$$\begin{aligned} (|v_1\rangle + |v_2\rangle) \otimes |w\rangle &= |v_1\rangle \otimes |w\rangle + |v_2\rangle \otimes |w\rangle, \\ |v\rangle \otimes (|w_1\rangle + |w_2\rangle) &= |v\rangle \otimes |w_1\rangle + |v\rangle \otimes |w_2\rangle, \\ \alpha(|v\rangle \otimes |w\rangle) &= (\alpha|v\rangle) \otimes |w\rangle = |v\rangle \otimes (\alpha|w\rangle), \end{aligned}$$

and equipped with the inner product given by

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

where $|v\rangle, |v_1\rangle, |v_2\rangle \in \mathcal{V}$, $|w\rangle, |w_1\rangle, |w_2\rangle \in \mathcal{W}$ and $\alpha \in \mathbb{C}$. \diamond

⁸ Note that the tensor product \otimes takes precedence over addition $+$ (and subtraction $-$).

REMARK 4.2. The notion of the tensor product can be naturally extended to finitely many Hilbert spaces, e.g. $\mathcal{U} \otimes \mathcal{V} \otimes \mathcal{W}$. This relies on the associativity of the tensor product: $(\mathcal{U} \otimes \mathcal{V}) \otimes \mathcal{W} = \mathcal{U} \otimes (\mathcal{V} \otimes \mathcal{W})$. \diamond

It is important to note that *not* all elements of $\mathcal{V} \otimes \mathcal{W}$ are of the form $|v\rangle \otimes |w\rangle$. Since elements of the form $|v\rangle \otimes |w\rangle$ *span* the Hilbert space $\mathcal{V} \otimes \mathcal{W}$, a general element is of the form

$$\sum_i \alpha_i |v_i\rangle \otimes |w_i\rangle$$

and the inner product between two general elements is

$$\left(\sum_i \overline{\alpha_i} \langle v_i| \otimes \langle w_i| \right) \left(\sum_j \beta_j |v'_j\rangle \otimes |w'_j\rangle \right) = \sum_{i,j} \overline{\alpha_i} \beta_j \langle v_i|v'_j\rangle \langle w_i|w'_j\rangle,$$

where $|v_i\rangle, |v'_i\rangle \in \mathcal{V}$, $|w_i\rangle, |w'_i\rangle \in \mathcal{W}$, and $\alpha_i, \beta_j \in \mathbb{C}$. We will return to this important problem later in this chapter when we discuss entanglement.

EXAMPLE 4.3. If $\mathcal{V} = \mathbb{C}^m$ and $\mathcal{W} = \mathbb{C}^n$, then $\mathcal{V} \otimes \mathcal{W} \simeq \mathbb{C}^{mn}$ with the usual inner product. More concretely, if (as a coordinate vector with respect to some basis)

$$\mathbb{C}^3 \ni |v\rangle = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \quad \mathbb{C}^2 \ni |w\rangle = \begin{pmatrix} 4 \\ 5 \end{pmatrix},$$

then

$$|v\rangle \otimes |w\rangle = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} \otimes \begin{pmatrix} 4 \\ 5 \end{pmatrix} \text{ can be identified with } \begin{pmatrix} 1 \cdot 4 \\ 1 \cdot 5 \\ 2 \cdot 4 \\ 2 \cdot 5 \\ 3 \cdot 4 \\ 3 \cdot 5 \end{pmatrix} = \begin{pmatrix} 4 \\ 5 \\ 8 \\ 10 \\ 12 \\ 15 \end{pmatrix} \in \mathbb{C}^6,$$

that is, we multiply each coordinate of the first vector with each coordinate of the second vector. \diamond

While many of the postulates of quantum theory seem rather ad hoc, if we accept the state postulate and require a bilinear map from the Hilbert spaces of two systems to the Hilbert space of the joint system, then the tensor product is essentially the only option. This is because of the following property of the tensor product, called the *universal property of the tensor product*, which will, however, not be of direct relevance to this module.



THEOREM 4.4. Let \mathcal{V}, \mathcal{W} be two vector spaces. The map

$$\varphi : \mathcal{V} \times \mathcal{W} \rightarrow \mathcal{V} \otimes \mathcal{W}, \quad (|v\rangle, |w\rangle) \mapsto |v\rangle \otimes |w\rangle$$

has the property that for any bilinear map $f : \mathcal{V} \times \mathcal{W} \rightarrow \mathcal{U}$ from $\mathcal{V} \times \mathcal{W}$ to another vector space \mathcal{U} there exists a unique linear map $g : \mathcal{V} \otimes \mathcal{W} \rightarrow \mathcal{U}$ such that $f = g \circ \varphi$. This fact is also expressed by saying that the diagram displayed on the right commutes.

$$\begin{array}{ccc} \mathcal{V} \times \mathcal{W} & \xrightarrow{\varphi} & \mathcal{V} \otimes \mathcal{W} \\ & \searrow f & \downarrow g \\ & & \mathcal{U} \end{array}$$

FIGURE 4.1. Commutative diagram for the universal property of the tensor product.

Moving back to practical issues, if we have basis on \mathcal{V} and a basis on \mathcal{W} , it is easy to construct a basis on $\mathcal{V} \otimes \mathcal{W}$:

PROPOSITION 4.5. *If $\{|v_1\rangle, |v_2\rangle, \dots, |v_m\rangle\}$ and $\{|w_1\rangle, |w_2\rangle, \dots, |w_n\rangle\}$ are bases for \mathcal{V} and \mathcal{W} , then*

$$\begin{aligned} \{|v_i\rangle \otimes |w_j\rangle\}_{i,j} = & \{|v_1\rangle \otimes |w_1\rangle, \dots, |v_1\rangle \otimes |w_n\rangle, \\ & |v_2\rangle \otimes |w_1\rangle, \dots, |v_2\rangle \otimes |w_n\rangle, \\ & \dots, \\ & |v_m\rangle \otimes |w_1\rangle, \dots, |v_m\rangle \otimes |w_n\rangle\} \end{aligned} \quad (4.1)$$

is a basis for $\mathcal{V} \otimes \mathcal{W}$, i.e., we take the tensor product of each pair of basis vectors. Moreover, if both original bases are orthonormal, also the resulting basis is.

PROOF. By definition, $\mathcal{V} \otimes \mathcal{W}$ is spanned by the vectors $|v\rangle \otimes |w\rangle$ with $|v\rangle \in \mathcal{V}$ and $|w\rangle \in \mathcal{W}$. Now we can (uniquely) decompose $|v\rangle$ and $|w\rangle$ in the bases $\{|v_1\rangle, |v_2\rangle, \dots, |v_m\rangle\}$ resp. $\{|w_1\rangle, |w_2\rangle, \dots, |w_n\rangle\}$:

$$|v\rangle = \sum_{i=1}^m \alpha_i |v_i\rangle, \quad |w\rangle = \sum_{j=1}^n \beta_j |w_j\rangle,$$

for some complex coefficients α_i, β_j . Together with the linearity of the tensor product this implies that $\mathcal{V} \otimes \mathcal{W}$ is spanned by the vectors (4.1). Linear independence in general can be shown using the universal property of the tensor product, but we will only look at the case of orthonormal bases. In particular, the orthonormality of the vectors $\{|v_i\rangle \otimes |w_j\rangle\}_{i,j}$, given that the bases for \mathcal{V} and \mathcal{W} are already orthonormal, follows from

$$(\langle v_i | \otimes \langle w_j |)(|v_k\rangle \otimes |w_l\rangle) = \langle v_i | v_k\rangle \cdot \langle w_j | w_l\rangle = \delta_{ik} \delta_{jl}.$$

It is then straightforward to show that any orthonormal (or just orthogonal) set is linearly independent. In summary, the set of vectors (4.1) is linearly independent and spans the space $\mathcal{V} \otimes \mathcal{W}$, and therefore it is a basis of $\mathcal{V} \otimes \mathcal{W}$. \square

COROLLARY 4.6. *We have*

$$\dim(\mathcal{V} \otimes \mathcal{W}) = \dim \mathcal{V} \cdot \dim \mathcal{W}.$$

That is, the dimension of the tensor product is the product of the dimensions.

COROLLARY 4.7. *Since every n -dimensional Hilbert space is isomorphic to \mathbb{C}^n (see Theorem 2.14), it follows from the above corollary that*

$$\mathbb{C}^n \otimes \mathbb{C}^m \simeq \mathbb{C}^{nm}.$$

EXAMPLE 4.8. Taking tensor products of the vectors in the $\{|0\rangle, |1\rangle\}$ basis of \mathbb{C}^2 , we obtain the basis

$$\{|0\rangle \otimes |0\rangle, |0\rangle \otimes |1\rangle, |1\rangle \otimes |0\rangle, |1\rangle \otimes |1\rangle\}$$

of $\mathbb{C}^2 \otimes \mathbb{C}^2 \simeq \mathbb{C}^4$. A general vector in $\mathbb{C}^2 \otimes \mathbb{C}^2$ is thus

$$\alpha_0 |0\rangle \otimes |0\rangle + \alpha_1 |0\rangle \otimes |1\rangle + \alpha_2 |1\rangle \otimes |0\rangle + \alpha_3 |1\rangle \otimes |1\rangle$$

for some $\alpha_0, \alpha_1, \alpha_2, \alpha_3 \in \mathbb{C}$. In the next section we will introduce a simplified notation for these tensor products. \diamond

Tensor products of Hilbert spaces play an important role in the composition of subsystems in quantum physics. For example, if $\mathcal{H}_1 = \mathbb{C}^2$ and $\mathcal{H}_2 = \mathbb{C}^2$ describe two single qubits, then $\mathcal{H}_1 \otimes \mathcal{H}_2$ describes the joint system of two qubits.

DEFINITION 4.9. An *n-qubit system* represented by the 2^n -dimensional Hilbert space

$$(\mathbb{C}^2)^{\otimes n} := \underbrace{\mathbb{C}^2 \otimes \cdots \otimes \mathbb{C}^2}_{n \text{ times}}. \quad \diamond$$

Note the exponential growth in the dimension of the Hilbert space. This is one of the reasons why quantum computers with comparatively few qubits could outperform even the largest classical computers in some tasks.

4.2 STATE SPACE

In this section, we take a closer look at the state space of n -qubit systems and introduce some useful notation.

TERMINOLOGY & NOTATION. Note that while people often refer to the set of quantum states (in a given dimension) as the *state space*, this is in fact not a vector space! It is just a set. Indeed, for two states $|\psi\rangle, |\varphi\rangle \in \mathcal{H}$, in general $|\psi\rangle + |\varphi\rangle$ is not normalised, and for any $\alpha \in \mathbb{C}$ such that $|\alpha| \neq 1$, the vector $\alpha|\psi\rangle$ is not normalised either. Nevertheless, we will sometimes use the terminology “state space”, as it often appears in the literature. \diamond

The *computational* or *standard basis* of a n -qubit system $(\mathbb{C}^2)^{\otimes n}$ is

$$\begin{aligned} &\{ |0\rangle \otimes \cdots \otimes |0\rangle \otimes |0\rangle, \\ &\quad |0\rangle \otimes \cdots \otimes |0\rangle \otimes |1\rangle, \\ &\quad |0\rangle \otimes \cdots \otimes |1\rangle \otimes |0\rangle, \\ &\quad \dots, \\ &\quad |1\rangle \otimes \cdots \otimes |1\rangle \otimes |1\rangle \}. \end{aligned}$$

Since such systems occur frequently, it is useful to write $|j_1 \cdots j_n\rangle$ instead of $|j_1\rangle \otimes \cdots \otimes |j_n\rangle$, so that this basis is alternatively written as

$$\{|0 \dots 00\rangle, |0 \dots 01\rangle, |0 \dots 10\rangle, \dots, |1 \dots 11\rangle\}.$$

Moreover, since decimal notation is much more compact than binary notation, we also write⁹

$$\{|0\rangle, |1\rangle, |2\rangle, \dots, |2^n - 1\rangle\}$$

⁹ For the decimal notation to be unambiguous, the number of qubits must be clear!

for the same vectors. Also, we will sometimes drop the tensor product notation, and use the notation $|\psi\rangle|\varphi\rangle$ instead of $|\psi\rangle \otimes |\varphi\rangle$.

For example, the computational basis for a two-qubit system $\mathbb{C}^2 \otimes \mathbb{C}^2$ has the equivalent notations

$$\begin{aligned} & \{|0\rangle \otimes |0\rangle, |0\rangle \otimes |1\rangle, |1\rangle \otimes |0\rangle, |1\rangle \otimes |1\rangle\} \\ &= \{|0\rangle|0\rangle, |0\rangle|1\rangle, |1\rangle|0\rangle, |1\rangle|1\rangle\} \\ &= \{|00\rangle, |01\rangle, |10\rangle, |11\rangle\} \\ &= \{|0\rangle, |1\rangle, |2\rangle, |3\rangle\}, \end{aligned}$$

while the computational basis for a three-qubit system $\mathbb{C}^2 \otimes \mathbb{C}^2 \otimes \mathbb{C}^2$ has the equivalent notations

$$\begin{aligned} & \{|0\rangle \otimes |0\rangle \otimes |0\rangle, |0\rangle \otimes |0\rangle \otimes |1\rangle, |0\rangle \otimes |1\rangle \otimes |0\rangle, \dots, |1\rangle \otimes |1\rangle \otimes |1\rangle\} \\ &= \{|0\rangle|0\rangle|0\rangle, |0\rangle|0\rangle|1\rangle, |0\rangle|1\rangle|0\rangle, \dots, |1\rangle|1\rangle|1\rangle\} \\ &= \{|000\rangle, |001\rangle, |010\rangle, \dots, |111\rangle\} \\ &= \{|0\rangle, |1\rangle, |2\rangle, |3\rangle, |4\rangle, |5\rangle, |6\rangle, |7\rangle\} \end{aligned}$$

Just as for single qubits, once an ordering of the basis vectors is fixed, we can also use coordinate vectors to represent vectors in $(\mathbb{C}^2)^{\otimes n}$. For $n = 2$ the standard ordering results in

$$|00\rangle = |0\rangle = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \quad |01\rangle = |1\rangle = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \quad |10\rangle = |2\rangle = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \quad |11\rangle = |3\rangle = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}.$$

It should be clear how this can be generalised to $n > 2$, but then the vectors become rather long—they have 2^n components.

For two-qubit systems the so-called *Bell states* will play a prominent role.

DEFINITION 4.11. The four *Bell states* are defined as

$$\begin{aligned} |\Phi^+\rangle &:= \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle), & |\Psi^+\rangle &:= \frac{1}{\sqrt{2}}(|01\rangle + |10\rangle), \\ |\Phi^-\rangle &:= \frac{1}{\sqrt{2}}(|00\rangle - |11\rangle), & |\Psi^-\rangle &:= \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle). \end{aligned} \quad \diamond$$

REMARK 4.12. It is straightforward to verify that the four Bell states form an orthonormal basis of the Hilbert space $\mathbb{C}^2 \otimes \mathbb{C}^2 \simeq \mathbb{C}^4$. \diamond

EXAMPLE 4.13. Here are two examples of state vectors for a three-qubit system, written in different notations:

$$\begin{aligned} \frac{1}{\sqrt{2}}(|0\rangle + |7\rangle) &= \frac{1}{\sqrt{2}}(|000\rangle + |111\rangle), \\ \frac{1}{2}(|1\rangle + |2\rangle + |4\rangle + |7\rangle) &= \frac{1}{2}(|001\rangle + |010\rangle + |100\rangle + |111\rangle). \end{aligned}$$

We check that they are indeed normalised:

$$\begin{aligned}
 & \frac{1}{2}(\langle 0| + \langle 7|)(|0\rangle + |7\rangle) \\
 &= \frac{1}{2}(\langle 0|0\rangle + \langle 0|7\rangle + \langle 7|0\rangle + \langle 7|7\rangle) = \frac{1}{2}(1 + 0 + 0 + 1) = 1, \\
 & \frac{1}{4}(\langle 1| + \langle 2| + \langle 4| + \langle 7|)(|1\rangle + |2\rangle + |4\rangle + |7\rangle) \\
 &= \frac{1}{4}(\langle 1|1\rangle + \langle 1|2\rangle + \cdots + \langle 7|4\rangle + \langle 7|7\rangle) = 1. \quad \diamond
 \end{aligned}$$

Just as for single qubits, unit vectors that differ only by a global phase represent the same state—global phases have no physical meaning.

EXAMPLE 4.14. Consider a two-qubit system and let $\phi \in \mathbb{R}$. The states $|00\rangle$ and $e^{i\phi}|00\rangle$ are equivalent. We write $|00\rangle \sim e^{i\phi}|00\rangle$. However, we have

$$\frac{1}{\sqrt{2}}(|00\rangle + |11\rangle) \approx \frac{1}{\sqrt{2}}(e^{i\phi}|00\rangle + |11\rangle).$$

In this second example, $e^{i\phi}$ is a relative phase, and it cannot be factored out, just as in the similar examples in the single-qubit case, e.g. $\frac{1}{\sqrt{2}}(|0\rangle + |1\rangle) \approx \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle)$. \diamond

Be cautious with the different but equivalent ways that a tensor product can be written. For example, the vectors

$$e^{i\phi}|0\rangle \otimes |0\rangle, \quad |0\rangle \otimes e^{i\phi}|0\rangle, \quad e^{i\phi}(|0\rangle \otimes |0\rangle)$$

do not only represent the same state, they are actually equal.

Another possible cause of confusion is the use of different bases to express a vector. For example, while the two vectors

$$\frac{1}{\sqrt{2}}(|+\rangle \otimes |+\rangle + |-\rangle \otimes |-\rangle) = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle).$$

look quite different, they are actually equal.

Finally, let's take another look at the number of parameters required to describe an n -qubit state up to a global phase:

PROPOSITION 4.15. *Every state for a n -qubit system can be uniquely represented in the computational basis as*

$$\alpha_0|0\rangle + \alpha_1|1\rangle + \cdots + \alpha_{2^n-1}|2^n-1\rangle, \quad (4.2)$$

for some complex coefficients α_i such that the first non-zero coefficient is positive.

PROOF. Every vector in $(\mathbb{C}^2)^{\otimes n}$ can be decomposed in the computational basis to obtain an expression of the form (4.2). Suppose that the first non-zero entry is α_i . Using the equivalence relation between state vectors, we can multiply the vector by $|\alpha_i|^{-1}\alpha_i$ to obtain an equivalent vector where the first non-zero entry is $|\alpha_i| > 0$. \square

Since the normalisation condition actually fixes the modulus of the first non-zero coefficient (we need $\sum_{i=0}^{2^n-1} |\alpha_i|^2 = 1$), the state space of an n -qubit system has complex dimension $2^n - 1$. Note again that this is exponential in the number of qubits n .

4.3 ENTANGLEMENT

Entanglement is one of the most fundamental and counter-intuitive features of quantum theory.¹⁰ It refers to the fact that quantum particles can be correlated to one another significantly stronger than classical particles can be correlated. Among other things, this strong correlation is what's behind quantum advantage in many quantum information processing protocols, and behind some peculiar features of quantum theory. In this section, we look at some basic definitions related to entanglement.

We have seen in section 3.1, that single-qubit states are completely specified (up to a global phase) by two real numbers (i.e. by their Bloch representation), that is, equivalently, by a single complex number. If we now take the tensor product of n single-qubit states, we obtain a state for an n -qubit system which is specified by n complex numbers. However, as we have seen in the previous section, the state space for n -qubit systems is vastly bigger. It has complex dimension $2^n - 1$ and thus we need $2^n - 1$ complex numbers to describe a general state of a n -qubit system. Consequently, most states cannot be described in terms of states of n single-qubit systems. When we look at composite quantum systems, *the whole is really greater than the sum (tensor product) of its parts.*

It is exactly these states that cannot be described as a simple “collection” (tensor product) of single-system states that will be interesting for us, and these are the *entangled* states. In full generality (not just for qubit states), entangled states are defined as follows.

DEFINITION 4.16. Let $\mathcal{H}_1, \dots, \mathcal{H}_n$ be Hilbert spaces and $|v\rangle \in \mathcal{H}_1 \otimes \dots \otimes \mathcal{H}_n$ a unit vector (a normalised vector). If there exist (unit) vectors $|v_i\rangle \in \mathcal{H}_i$, $i = 1, \dots, n$, such that

$$|v\rangle = |v_1\rangle \otimes \dots \otimes |v_n\rangle,$$

then $|v\rangle$ is called a *product state*. Otherwise it is called an *entangled state*. \diamond

As we have seen above, the state space of an n -qubit system is $2^n - 1$ dimensional, while the space of single-qubit product states is only a $n \ll 2^n - 1$ dimensional subspace (*subset*, to be more precise). Therefore, essentially every state is entangled.

EXAMPLE 4.17. The Bell states are entangled. For instance, $|\Phi^+\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$ cannot be decomposed as

$$(\alpha_1 |0\rangle + \beta_1 |1\rangle) \otimes (\alpha_2 |0\rangle + \beta_2 |1\rangle) = \alpha_1 \alpha_2 |00\rangle + \alpha_1 \beta_2 |01\rangle + \alpha_2 \beta_1 |10\rangle + \beta_1 \beta_2 |11\rangle$$

because $\alpha_1 \beta_2 = 0$ implies that either $\alpha_1 \alpha_2 = 0$ or $\beta_1 \beta_2 = 0$. \diamond

¹⁰ So much so that Einstein, Podolsky and Rosen wrote an [article](#) about how some consequences of entanglement should mean that quantum theory is an incomplete theory. So far, nobody has come up with a better theory, so perhaps we just need to accept entanglement.

While the notion of entanglement is completely *basis independent* (different than, say, superposition), it is always with respect to a specified tensor product decomposition.

EXAMPLE 4.18. The three-qubit state

$$|v\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |3\rangle) = \frac{1}{\sqrt{2}}(|000\rangle + |011\rangle)$$

is entangled with respect to the single-qubit decomposition $\mathbb{C}^2 \otimes \mathbb{C}^2 \otimes \mathbb{C}^2$, because there exist no $\alpha_i, \beta_i \in \mathbb{C}$, $i = 1, 2, 3$, such that

$$|v\rangle = (\alpha_1 |0\rangle + \beta_1 |1\rangle) \otimes (\alpha_2 |0\rangle + \beta_2 |1\rangle) \otimes (\alpha_3 |0\rangle + \beta_3 |1\rangle).$$

It is unentangled, however, with respect to the decomposition $\mathbb{C}^2 \otimes \mathcal{H}$, with $\mathcal{H} = \mathbb{C}^2 \otimes \mathbb{C}^2$, into a single- and a two-qubit system, because

$$|v\rangle = |0\rangle \otimes \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle) = |0\rangle \otimes |\Phi^+\rangle$$

is a product state with respect to this decomposition ($|0\rangle$ and $|\Phi^+\rangle$ are unit vectors in \mathbb{C}^2 resp. \mathcal{H}). \diamond

In the bipartite case (two systems), the description of states on the Hilbert space $\mathbb{C}^d \otimes \mathbb{C}^d$ requires $d^2 - 1$ complex numbers. However, in this case, there always exist two bases on \mathbb{C}^d (one for the first tensor factor and one for the second) such that in these bases, the description is much more compact, requiring only $d - 1$ real numbers.

THEOREM 4.19 (SCHMIDT DECOMPOSITION). *Consider a quantum state $|\psi\rangle \in \mathbb{C}^d \otimes \mathbb{C}^d$. Then, there exist two orthonormal bases on \mathbb{C}^d , $\{|e_i\rangle\}_{i=0}^{d-1}$ and $\{|f_j\rangle\}_{j=0}^{d-1}$ such that*

$$|\psi\rangle = \sum_{i=0}^{d-1} \lambda_i |e_i\rangle \otimes |f_i\rangle,$$

where $\lambda_i \geq 0$ and $\sum_{i=0}^{d-1} \lambda_i^2 = 1$. This is called the Schmidt decomposition of the state $|\psi\rangle$.

PROOF. Let us write $|\psi\rangle$ in the computational basis,

$$|\psi\rangle = \sum_{i=0}^{d-1} \sum_{j=0}^{d-1} a_{ij} |i\rangle \otimes |j\rangle. \quad (4.3)$$

Consider the $d \times d$ matrix A with elements a_{ij} . By the singular value decomposition, A can be written as

$$A = U \Lambda V^\dagger,$$

where U and V are unitary matrices and Λ is a diagonal matrix with non-negative entries, i.e. $(\Lambda)_{ij} = \delta_{ij} \lambda_i$ for some $\lambda_i \geq 0$. That is, we have

$$a_{ij} = \sum_{k=0}^{d-1} \sum_{\ell=0}^{d-1} (U)_{ik} \Lambda_{k\ell} (V^\dagger)_{\ell j} = \sum_{k=0}^{d-1} u_{ik} \lambda_k \bar{v}_{jk},$$

where we used that $(V^\dagger)_{ij} = \overline{(V)_{ji}} =: \bar{v}_{ji}$. Substituting this into Equation (4.3), we get

$$|\psi\rangle = \sum_{i=0}^{d-1} \sum_{j=0}^{d-1} \sum_{k=0}^{d-1} u_{ik} \lambda_k \bar{v}_{jk} |i\rangle \otimes |j\rangle = \sum_{k=0}^{d-1} \lambda_k \left(\sum_{i=0}^{d-1} u_{ik} |i\rangle \right) \otimes \left(\sum_{j=0}^{d-1} \bar{v}_{jk} |j\rangle \right).$$

We can now define $|e_k\rangle = \sum_{i=0}^{d-1} u_{ik} |i\rangle$ and $|f_k\rangle = \sum_{j=0}^{d-1} \bar{v}_{jk} |j\rangle$. From the unitarity of U and V , it follows that $\{|e_k\rangle\}$ and $\{|f_k\rangle\}$ are orthonormal bases, and substituting these into the above equation proves the theorem. \square

The Schmidt decomposition is quite useful in characterising entanglement, or simply for efficiently representing entangled states.

4.4 OPERATORS

Apart from the structure of states on tensor product Hilbert spaces, it is also important to look at operators, as these describe measurements and time evolution of composite systems. First, we introduce the notion of the tensor product of operators.

DEFINITION 4.20. Suppose that A and B are operators on the Hilbert spaces \mathcal{V} and \mathcal{W} , respectively. Then we define on $\mathcal{V} \otimes \mathcal{W}$ an operator $A \otimes B$, called the tensor product of A and B , by setting, for any $|v\rangle \in \mathcal{V}$ and $|w\rangle \in \mathcal{W}$,

$$(A \otimes B)(|v\rangle \otimes |w\rangle) := (A|v\rangle) \otimes (B|w\rangle)$$

and extending this definition linearly to general vectors in $\mathcal{V} \otimes \mathcal{W}$. \diamond

EXAMPLE 4.21. Consider, for example, a two-qubit system in the state $|00\rangle$. Applying the operator $X \otimes H$ to it yields

$$(X \otimes H)|00\rangle = (X|0\rangle) \otimes (H|0\rangle) = |1\rangle \otimes |+\rangle = \frac{1}{\sqrt{2}}(|10\rangle + |11\rangle). \quad \diamond$$

Applying the definition above twice, we note that if C and D are also operators on \mathcal{V} and \mathcal{W} , respectively, then

$$(A \otimes B)(C \otimes D) = (AC) \otimes (BD).$$

In particular, we have

$$(A \otimes \mathbf{1})(\mathbf{1} \otimes B) = A \otimes B,$$

where $\mathbf{1}$ denotes the identity operator on any Hilbert space.

The important notions of the adjoint and the inverse carry through the tensor product:

PROPOSITION 4.22. *For any two operators A and B on Hilbert spaces \mathcal{V} and \mathcal{W} , respectively, we have*

$$(A \otimes B)^\dagger = A^\dagger \otimes B^\dagger.$$

Moreover, if A and B are both invertible, we have

$$(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}.$$

PROOF. Let $|v\rangle, |v'\rangle \in \mathcal{V}$ and $|w\rangle, |w'\rangle \in \mathcal{W}$ be arbitrary. Then

$$\begin{aligned} (\langle v| \otimes \langle w|) (A \otimes B)^\dagger (|v'\rangle \otimes |w'\rangle) &= \overline{(\langle v'| \otimes \langle w'|) (A \otimes B) (|v\rangle \otimes |w\rangle)} \\ &= \overline{\langle v'| A | v \rangle \langle w'| B | w \rangle} \\ &= \langle v| A^\dagger | v' \rangle \langle w| B^\dagger | w' \rangle \\ &= (\langle v| \otimes \langle w|) (A^\dagger \otimes B^\dagger) (|v'\rangle \otimes |w'\rangle). \end{aligned}$$

Since every vector in $\mathcal{V} \otimes \mathcal{W}$ can be written as a sum of product states, this implies $(A \otimes B)^\dagger = A^\dagger \otimes B^\dagger$ by linearity. If A, B are invertible, then $A^{-1} \otimes B^{-1}$ satisfies

$$\mathbb{1} \otimes \mathbb{1} = (A \otimes B) (A^{-1} \otimes B^{-1})$$

and thus is the inverse of $A \otimes B$, denoted $(A \otimes B)^{-1}$. \square

COROLLARY 4.23. If A, B are unitary/self-adjoint operators on \mathcal{V} resp. \mathcal{W} , then $A \otimes B$ is a unitary/self-adjoint operator on $\mathcal{V} \otimes \mathcal{W}$. \square

EXAMPLE 4.24. It follows that the two-qubit operator $X \otimes H$ is unitary. Other examples of such unitary operators are $X \otimes \mathbb{1}$ or $H \otimes H$. \diamond

As for states, it is not the case that all operators on a tensor product Hilbert space can be constructed by a single tensor product of operators on its factors. That is, in general there are no operators A and B on \mathcal{V} resp. \mathcal{W} such that a generic operator C on $\mathcal{V} \otimes \mathcal{W}$ can be written as $C = A \otimes B$. However—again, as for states—the vector space of operators on a (finite-dimensional) tensor product Hilbert space is *spanned* by tensor products of operators on its factors. That is, every operator C on the tensor product $\mathcal{V} \otimes \mathcal{W}$ can be constructed by (finite) sums of tensor products of operators,

$$C = \sum_i A_i \otimes B_i$$

for some operators A_i on \mathcal{V} and B_i on \mathcal{W} .

Let us look at an important example: the *controlled-NOT* or *CNOT* gate is the operator on $\mathbb{C}^2 \otimes \mathbb{C}^2$ defined by

$$\text{CNOT} := C_X := |00\rangle\langle 00| + |01\rangle\langle 01| + |11\rangle\langle 10| + |10\rangle\langle 11|.$$

One can check that this operator is unitary. We use the notation C_X because of its relation to the X gate,

$$C_X = |0\rangle\langle 0| \otimes \mathbb{1} + |1\rangle\langle 1| \otimes X,$$

Therefore it is also sometimes called the *controlled- X gate*. In a similar way we can turn any single-qubit gate, say U , into a controlled gate C_U .

To see why we use the term *controlled*, consider the action of C_X on a general two-qubit state

$$|v\rangle = \alpha_0 |00\rangle + \alpha_1 |01\rangle + \alpha_2 |10\rangle + \alpha_3 |11\rangle.$$

The action is given by

$$C_X |v\rangle = \alpha_0 |00\rangle + \alpha_1 |01\rangle + \alpha_2 |11\rangle + \alpha_3 |10\rangle,$$

that is, C_X swaps the basis vectors $|10\rangle$ and $|11\rangle$ and leaves $|00\rangle$ and $|01\rangle$ invariant. That is, the CNOT gate flips the second qubit (*target qubit*) if and only if the first qubit (*control qubit*) is in state $|1\rangle$, otherwise it leaves the target qubit invariant. But since we are dealing with quantum states, we can use this gate to do some interesting things, such as creating entanglement, by e.g. preparing the control qubit in state $|+\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$.

EXAMPLE 4.25. A combination of the Hadamard gate H and the CNOT gate can be used to obtain the (entangled) Bell state $|\Phi^+\rangle$ from the (product) state $|00\rangle$. Namely,

$$C_X (H \otimes \mathbb{1}) |00\rangle = C_X |+\rangle |0\rangle = \frac{1}{\sqrt{2}} C_X (|00\rangle + |10\rangle) = \frac{1}{\sqrt{2}} (|00\rangle + |11\rangle) = |\Phi^+\rangle. \quad \diamond$$

In the computational basis with standard ordering, the CNOT gate can be written as

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

Let us also introduce the *SWAP gate*. It is the operator on $\mathbb{C}^2 \otimes \mathbb{C}^2$ defined by

$$\text{SWAP} = |00\rangle\langle 00| + |01\rangle\langle 10| + |10\rangle\langle 01| + |11\rangle\langle 11|.$$

As the name suggests, it interchanges the states of a product state. Suppose that

$$|\psi\rangle = \alpha |0\rangle + \beta |1\rangle, \quad |\varphi\rangle = \gamma |0\rangle + \delta |1\rangle$$

are two general single-qubit states. Then

$$\begin{aligned} \text{SWAP} (|\psi\rangle \otimes |\varphi\rangle) &= \text{SWAP} (\alpha \gamma |00\rangle + \alpha \delta |01\rangle + \beta \gamma |10\rangle + \beta \delta |11\rangle) \\ &= \gamma \alpha |00\rangle + \gamma \beta |01\rangle + \delta \alpha |10\rangle + \delta \beta |11\rangle \\ &= |\varphi\rangle \otimes |\psi\rangle. \end{aligned}$$

While not strictly part of this module, the SWAP operator plays an important role in characterising entanglement:



PROPOSITION 4.26. A unitary operator U on $\mathbb{C}^2 \otimes \mathbb{C}^2$ maps all product states to product states (and all entangled states to entangled states) if and only if it is of the form

$$U_1 \otimes U_2 \quad \text{or} \quad \text{SWAP} (U_1 \otimes U_2)$$

for unitaries U_1, U_2 on \mathbb{C}^2 .

4.5 MEASUREMENTS

In Section 2.2, we saw how a measurement of a single system changes the state of that system in a probabilistic way dependent on the measuring device.

With multiple systems we can also measure in any orthonormal basis. A physically important example are the bases given by Proposition 4.5. These can be implemented by performing local measurements on each qubit. For example, suppose we have a two-qubit system with the first qubit held by Alice and the second qubit held by Bob. If Alice measures her qubit in the $\{|v_1\rangle, |v_2\rangle\}$ basis and Bob measures his qubit in the $\{|w_1\rangle, |w_2\rangle\}$ basis, we can describe this as a two-qubit measurement in the $\{|v_i\rangle \otimes |w_j\rangle\}_{i,j}$ basis. If the overall outcome is $|v_i\rangle \otimes |w_j\rangle$, we interpret this as Alice getting the $|v_i\rangle$ outcome and Bob getting the $|w_j\rangle$ outcome.

EXAMPLE 4.27. Consider each qubit of a two-qubit system being measured in the computational basis. This amounts to a measurement in the $\{|00\rangle, |01\rangle, |10\rangle, |11\rangle\}$ basis.

When such a device measures a general two-qubit state, $|\psi\rangle = \alpha_{00}|00\rangle + \alpha_{01}|01\rangle + \alpha_{10}|10\rangle + \alpha_{11}|11\rangle$, the probability of outcome $|ij\rangle$ is $|\alpha_{ij}|^2$. For example, such a measurement on the Bell state $|\Phi^+\rangle$ results in $|00\rangle$ and $|11\rangle$ with equal probability $\frac{1}{2}$. Suppose we only want to know the probability of the first qubit getting result $|0\rangle$. Since this occurs when the overall result is $|00\rangle$ or $|01\rangle$, the probability is $|\alpha_{00}|^2 + |\alpha_{01}|^2$. \diamond

Note something “spooky” in the above example: when Alice measures her qubit and obtains outcome $|0\rangle$, she *knows* that Bob’s outcome on the other qubit will also be $|0\rangle$, and similarly for the case of $|1\rangle$. In the language of post-measurement states, the two-qubit state (immediately!) after Alice observing outcome $|0\rangle$ is given by $|00\rangle$. This may sound like something that breaks the theory of relativity: Alice and Bob might be very far apart, but once Alice (locally) measures her state, the state on Bob’s side immediately gets updated. This weirdness is what triggered Einstein, Podolsky and Rosen to write a paper about how quantum theory cannot be a complete theory of physics. However, as we will see in a later chapter, this immediate update of the quantum state does *not* allow for faster-than-light communication, i.e. for *information* travelling from Alice to Bob faster than the speed of light (let alone immediately).

EXAMPLE 4.28. Similar to the last example, but now each qubit is measured in the Hadamard basis. The outcomes are $\{|+\rangle \otimes |+\rangle, |+\rangle \otimes |-\rangle, |-\rangle \otimes |+\rangle, |-\rangle \otimes |-\rangle\}$. Similar to above, such a measurement on the Bell state $|\Phi^+\rangle$ results in $|+\rangle \otimes |+\rangle$ and $|-\rangle \otimes |-\rangle$ with equal probability. \diamond

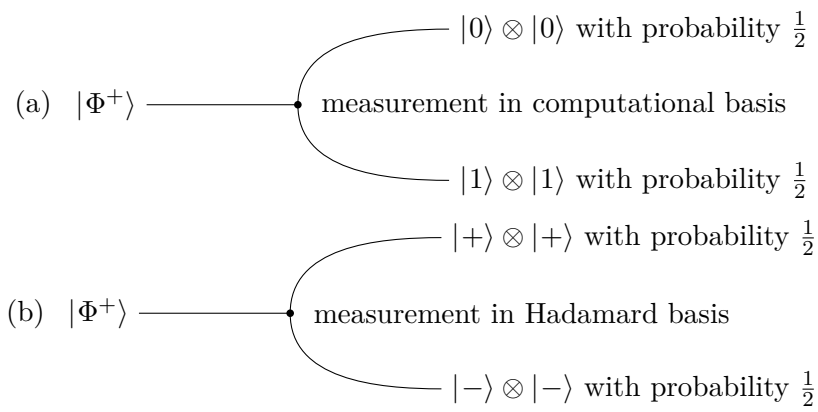


FIGURE 4.2. Illustrations corresponding to (a) Example 4.27 and (b) Example 4.28 with initial state $|\Phi^+\rangle$.

5 Density operators

Up to this point, we have been describing the state of a quantum system with a normalised element of a Hilbert space \mathcal{H} , in accordance with the state postulate. This postulate works well together with the other ones, such as the time evolution postulate, that describes the time evolution of a *closed* system by a unitary operator. However, we saw some examples where the system is not closed. This has consequences not just for time evolution, but also for the description of a quantum state of a non-closed system after it has undergone some time evolution (even a unitary one). In Example 4.25, we saw that the CNOT gate maps the product state $|+\rangle \otimes |0\rangle$ to the entangled Bell state $|\phi^+\rangle$. If we consider a single qubit of this two-qubit system, the single-qubit systems are not closed: they interact with each other through the CNOT gate. This leads to the problem that while before this interaction it is clear how to describe the state of the two qubits ($|+\rangle$ and $|0\rangle$), after the interaction we cannot give such a description of the individual qubit systems in terms of Hilbert space vectors, because $|\phi^+\rangle$ cannot be written as a tensor product of two qubit vectors.

Similarly, there are some physical situations in which the description of a quantum state with a Hilbert space vector is not quite adequate. Think of the BB84 protocol (Section 3.2) from the perspective of the eavesdropper, Eve. The state she receives is either one of the set $\{|0\rangle, |1\rangle, |+\rangle, |-\rangle\}$ with equal probabilities. She has no information whatsoever on which of these states was prepared. As such, a more physical description of the state she obtains (from her perspective) would be a probabilistic mixture of these four states.

The concept of *density operators* (or *density matrices*) helps with both of the above problems, and provides a more general description of quantum states than the Hilbert space vectors.

5.1 QUANTUM STATES AS OPERATORS

We begin by first providing a description of quantum states that is equivalent to the Hilbert space vector definition. Recall that the state postulate says that a quantum state is a normalised element $|\psi\rangle \in \mathcal{H}$, and two states $|\psi\rangle, |\varphi\rangle \in \mathcal{H}$ are equivalent if $|\varphi\rangle = e^{i\theta} |\psi\rangle$. We will associate to such a vector the following *density operator*:

$$|\psi\rangle \rightarrow |\psi\rangle\langle\psi|. \quad (5.1)$$

Note that if $|\psi\rangle \in \mathcal{H}$ then $|\psi\rangle\langle\psi| \in \mathcal{L}(\mathcal{H})$, and $|\psi\rangle\langle\psi|$ is a rank-1 projection. Importantly, the phase equivalence ($|\varphi\rangle = e^{i\theta}|\psi\rangle \sim |\psi\rangle$) is automatically encoded in the density operator. Indeed, the density operators corresponding to $|\varphi\rangle$ and $|\psi\rangle$ are not just equivalent, but they are *the same*:

$$|\varphi\rangle\langle\varphi| = (e^{i\theta}|\psi\rangle)(e^{i\theta}|\psi\rangle)^\dagger = e^{i\theta}|\psi\rangle\langle\psi|e^{-i\theta} = |\psi\rangle\langle\psi|.$$

EXAMPLE 5.1. Consider the density operator corresponding to the state $|0\rangle$, given in its matrix representation as

$$|0\rangle\langle 0| = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}. \quad \diamond$$

Re-defining the state (and therefore at least nominally changing the state postulate) according to Equation (5.1) can be justified by verifying that the rest of the postulates can also be equivalently re-stated in terms of the density operator instead of the Hilbert space vector. Before we do so, we revisit the definition of the trace of an operator.

DEFINITION 5.2. The *trace* of an operator $A \in \mathcal{L}(\mathcal{H})$ is given by

$$\text{tr}(A) = \sum_{j=1}^{\dim \mathcal{H}} \langle e_j | A | e_j \rangle,$$

where $\{|e_j\rangle\}_{j=1}^{\dim \mathcal{H}}$ is an *arbitrary* orthonormal basis of \mathcal{H} . \diamond

REMARK 5.3. Note that the definition of the trace is independent of the basis. An important property of the trace is *linearity*, that is, for all $A, B \in \mathcal{L}(\mathcal{H})$ and for all $\alpha, \beta \in \mathbb{C}$ we have

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

Another important property is *cyclicity*: for any $A, B \in \mathcal{L}(\mathcal{H})$ we have

$$\text{tr}(AB) = \text{tr}(BA).$$

In fact, cyclicity holds even if $A \in \mathcal{L}(\mathcal{H}, \mathcal{K})$ and $B \in \mathcal{L}(\mathcal{K}, \mathcal{H})$. Note that in this case $AB \in \mathcal{L}(\mathcal{K})$ and $BA \in \mathcal{L}(\mathcal{H})$.

Last, we have that for tensor products of operators, $A \in \mathcal{L}(\mathcal{H})$ and $B \in \mathcal{L}(\mathcal{K})$,

$$\text{tr}(A \otimes B) = \text{tr}(A) \text{tr}(B). \quad \diamond$$

The measurement postulate prescribes a set of n operators $\{M_j\}_{j=1}^n \subset \mathcal{L}(\mathcal{H})$ representing an n -outcome measurement, such that $\sum_{j=1}^n M_j^\dagger M_j = \mathbb{1}$. The probability of observing outcome j if the state is $|\psi\rangle$ is given by

$$p(j) = \langle \psi | M_j^\dagger M_j | \psi \rangle = \text{tr}(M_j^\dagger M_j |\psi\rangle\langle\psi|),$$

where the last equality can be seen either by picking a basis of \mathcal{H} that contains $|\psi\rangle$, or by using the cyclicity of the trace and recalling that kets

and bras are operators, $|\psi\rangle : \mathbb{C} \rightarrow \mathcal{H}$ and $\langle\psi| : \mathcal{H} \rightarrow \mathbb{C}$. We can therefore equivalently write the measurement probabilities in terms of the density operator $|\psi\rangle\langle\psi|$.

The post-measurement state after observing outcome j —according to the measurement postulate—is given by

$$|\varphi_j\rangle = \frac{M_j |\psi\rangle}{\sqrt{\langle\psi|M_j^\dagger M_j|\psi\rangle}} = \frac{M_j |\psi\rangle}{\sqrt{p(j)}}.$$

This can also be re-stated in terms of density operators, by assigning the following density operator to the post-measurement state:

$$\begin{aligned} |\varphi_j\rangle\langle\varphi_j| &= \left(\frac{M_j |\psi\rangle}{\sqrt{\langle\psi|M_j^\dagger M_j|\psi\rangle}} \right) \left(\frac{M_j |\psi\rangle}{\sqrt{\langle\psi|M_j^\dagger M_j|\psi\rangle}} \right)^\dagger = \frac{M_j |\psi\rangle\langle\psi|M_j^\dagger}{\langle\psi|M_j^\dagger M_j|\psi\rangle} \\ &= \frac{M_j |\psi\rangle\langle\psi|M_j^\dagger}{\text{tr}(M_j^\dagger M_j |\psi\rangle\langle\psi|)} = \frac{M_j |\psi\rangle\langle\psi|M_j^\dagger}{p(j)}. \end{aligned}$$

Therefore, we re-stated the measurement postulate in terms of density operators. Note that if we only care about outcome probabilities (and not the post-measurement states) and describe measurements by POVMs, then this can also be done by replacing $M_j^\dagger M_j$ by a positive semidefinite operator $N_j \geq 0$ in the above formulae. That is, a POVM is described by a set of n positive semidefinite operators, $\{N_j\}_{j=1}^n$ such that $\sum_j N_j = \mathbb{1}$, and the probability of outcome j on a state $|\psi\rangle\langle\psi|$ is given by

$$p(j) = \text{tr}(N_j |\psi\rangle\langle\psi|).$$

After the measurement postulate, we move on to re-stating the time evolution postulate in terms of density operators. Recall that the evolution of a (closed) system in terms of Hilbert space vectors is given by a unitary operation,

$$|\psi\rangle \rightarrow U |\psi\rangle.$$

We can therefore assign the corresponding density operator to the updated state instead:

$$|\psi\rangle\langle\psi| \rightarrow (U |\psi\rangle)(U |\psi\rangle)^\dagger = U |\psi\rangle\langle\psi| U^\dagger,$$

which leads to the time evolution postulate in terms of density operators.

REMARK 5.4. Note that the time evolution postulate assumed that the system is closed. In the case of non-closed systems (that interact with their environments), not just the description of the state, but also that of the time evolution needs to be changed in general. We will only cover this generalisation of the time evolution in the M-level material in this module. \diamond

The last remaining postulate concerns composite systems. This can also be re-stated in terms of density operators. Given a system associated to the Hilbert space \mathcal{H} , and another one associated to \mathcal{K} , the Hilbert space of the

joint system is given by $\mathcal{H} \otimes \mathcal{K}$. Clearly, we can define density operators on the tensor product Hilbert space. In the case of a product state, $|\psi\rangle \otimes |\varphi\rangle \in \mathcal{H} \otimes \mathcal{K}$, where $|\psi\rangle \in \mathcal{H}$ and $|\varphi\rangle \in \mathcal{K}$, we can define the density operator

$$(|\psi\rangle \otimes |\varphi\rangle)(\langle\psi| \otimes \langle\varphi|)^\dagger = (|\psi\rangle \otimes |\varphi\rangle)(\langle\psi| \otimes \langle\varphi|) = |\psi\rangle\langle\psi| \otimes |\varphi\rangle\langle\varphi|$$

on $\mathcal{H} \otimes \mathcal{K}$. Similarly, for a generic (perhaps entangled) state vector $|\psi\rangle \in \mathcal{H} \otimes \mathcal{K}$, we can define the corresponding density operator $|\psi\rangle\langle\psi| \in \mathcal{L}(\mathcal{H} \otimes \mathcal{K})$. We will return to the question of entanglement in terms of density operators soon, but for now, we see that all of our postulates can be re-phrased in terms of density operators corresponding to Hilbert space vectors.

5.2 PROBABILISTIC MIXTURES OF QUANTUM STATES

Thus far, the formulation of the postulates in terms of a density operator corresponding to a Hilbert space vector was just that: a re-formulation. However, working with density operators allows us to define a natural notion of a *probabilistic mixture* of quantum states. Imagine that some source prepares a quantum state $|\psi\rangle\langle\psi| \in \mathcal{L}(\mathcal{H})$ with probability q , and some other quantum state $|\varphi\rangle\langle\varphi| \in \mathcal{L}(\mathcal{H})$ with probability $1 - q$. The resulting state is a probabilistic mixture of the two states, and defines a valid physical process. As such, we should associate some mathematical object to it. It should be a quantum state, and therefore the measurement, evolution, and composition postulates should be possible to define on it. Furthermore, when for example we perform a measurement on this mixture, we expect the outcome distribution to also be a probabilistic mixture of the outcome distribution corresponding to $|\psi\rangle\langle\psi|$ and to $|\varphi\rangle\langle\varphi|$. The density operator formalism allows for a quite natural definition of this probabilistic mixture: the above described state is associated to the operator

$$\rho := q|\psi\rangle\langle\psi| + (1 - q)|\varphi\rangle\langle\varphi|, \quad (5.2)$$

which is again an operator on $\mathcal{L}(\mathcal{H})$.

We can now treat this operator ρ just like it was a quantum state, and apply our postulates. For example, consider a POVM described by $\{N_j\}_{j=1}^n$. The outcome probabilities on the states $|\psi\rangle\langle\psi|$ and $|\varphi\rangle\langle\varphi|$ are given by

$$p_\psi(j) = \text{tr}(N_j |\psi\rangle\langle\psi|), \quad \text{and} \quad p_\varphi(j) = \text{tr}(N_j |\varphi\rangle\langle\varphi|).$$

Now if we treat ρ in Equation (5.2) as a quantum state, we obtain the outcome probabilities

$$\begin{aligned} p_\rho(j) &= \text{tr}(N_j \rho) = \text{tr} \left[N_j (q|\psi\rangle\langle\psi| + (1 - q)|\varphi\rangle\langle\varphi|) \right] \\ &= q \text{tr}(N_j |\psi\rangle\langle\psi|) + (1 - q) \text{tr}(N_j |\varphi\rangle\langle\varphi|) = qp_\psi(j) + (1 - q)p_\varphi(j) \end{aligned}$$

as expected. Similarly, the time evolution of a probabilistic mixture is the probabilistic mixture of time evolutions:

$$U\rho U^\dagger = U \left[q|\psi\rangle\langle\psi| + (1 - q)|\varphi\rangle\langle\varphi| \right] U^\dagger = qU|\psi\rangle\langle\psi|U^\dagger + (1 - q)U|\varphi\rangle\langle\varphi|U^\dagger.$$

That is, the operator in Equation (5.2) indeed defines the physical situation of preparing the state $|\psi\rangle\langle\psi|$ with probability q and the state $|\varphi\rangle\langle\varphi|$ with probability $1 - q$.

REMARK 5.5. Note that the superposition $|\eta\rangle = \sqrt{q}|\psi\rangle + \sqrt{1-q}|\varphi\rangle$ does not satisfy these physical conditions, and therefore does not correspond to a probabilistic mixture. For example, the outcome probabilities for a POVM $\{N_j\}_{j=1}^n$ are given by

$$\begin{aligned} p_\eta(j) &= \langle\eta|N_j|\eta\rangle = (\sqrt{q}\langle\psi| + \sqrt{1-q}\langle\varphi|)N_j(\sqrt{q}|\psi\rangle + \sqrt{1-q}|\varphi\rangle) \\ &= q\langle\psi|N_j|\psi\rangle + (1-q)\langle\varphi|N_j|\varphi\rangle \\ &\quad + q(1-q)\langle\psi|N_j|\varphi\rangle + (1-q)q\langle\varphi|N_j|\psi\rangle \\ &\neq qp_\psi(j) + (1-q)p_\varphi(j). \end{aligned} \quad \diamond$$

EXAMPLE 5.6. Consider an equal mixture of the states $|0\rangle$ and $|1\rangle$, leading to the density operator

$$\frac{1}{2}|0\rangle\langle 0| + \frac{1}{2}|1\rangle\langle 1| = \frac{\mathbb{1}}{2},$$

which is often called the *maximally mixed state*. This name can be justified e.g. by noting that measuring this state with *any* POVM $\{M_j\}_{j=1}^n$ leads to the probabilities

$$p(j) = \text{tr}(M_j \frac{\mathbb{1}}{2}) = \frac{1}{2} \text{tr}(M_j),$$

which is simply a property of the POVM. That is, the state $\frac{\mathbb{1}}{2}$ does not convey any information that was not already present in the POVM itself. \diamond

Of course, one might consider the probabilistic mixture of more than two states. The mixture of the states $\{|\psi_j\rangle\}_{j=1}^n$ with respective probabilities $q_j \geq 0$ and $\sum_{j=1}^n q_j = 1$ is described by the density operator

$$\rho = \sum_{j=1}^n q_j |\psi_j\rangle\langle\psi_j|.$$

The following theorem provides a useful characterisation of all such density operators.

THEOREM 5.7. *The following are equivalent:*

1. An operator $\rho \in \mathcal{L}(\mathcal{H})$ is a density operator corresponding to some mixture of states, that is, $\rho = \sum_{j=1}^n q_j |\psi_j\rangle\langle\psi_j|$, where $|\psi_j\rangle \in \mathcal{H}$ are normalised vectors and $\{q_j\}$ is a probability distribution.
2. ρ is a positive semidefinite operator with trace one, that is, $\rho \geq 0$ and $\text{tr}(\rho) = 1$.

PROOF. 1. \implies 2.: To prove that ρ is positive semidefinite, we note that for all $|\varphi\rangle \in \mathcal{H}$ we have

$$\langle \varphi | \rho | \varphi \rangle = \langle \varphi | \sum_{j=1}^n q_j |\psi_j\rangle \langle \psi_j| | \varphi \rangle = \sum_{j=1}^n q_j \langle \varphi | \psi \rangle \langle \psi | \varphi \rangle = \sum_{j=1}^n q_j |\langle \varphi | \psi \rangle|^2 \geq 0,$$

since $q_j \geq 0$ for all j . To prove the trace property, direct calculation shows

$$\text{tr}(\rho) = \text{tr} \left(\sum_{j=1}^n q_j |\psi_j\rangle \langle \psi_j| \right) = \sum_{j=1}^n q_j \text{tr}(|\psi_j\rangle \langle \psi_j|) = \sum_{j=1}^n q_j = 1,$$

where we used that $\{q_j\}$ is a probability distribution and that $\text{tr}(|\psi_j\rangle \langle \psi_j|) = \langle \psi_j | \psi_j \rangle = 1$ for all j .

2. \implies 1.: To prove the converse direction, recall that $\rho \geq 0$ implies that ρ is Hermitian (Proposition 2.47), and therefore the spectral decomposition (Theorem 2.42) applies. In particular, ρ can be written as

$$\rho = \sum_{j=1}^{\dim \mathcal{H}} \lambda_j |e_j\rangle \langle e_j|, \quad (5.3)$$

where $\{|e_j\rangle\}_{j=1}^{\dim \mathcal{H}}$ is an orthonormal basis of \mathcal{H} and $\rho |e_j\rangle = \lambda_j |e_j\rangle$. The positivity condition implies that for all j , we have

$$\begin{aligned} 0 \leq \langle e_j | \rho | e_j \rangle &= \langle e_j | \sum_{k=1}^{\dim \mathcal{H}} \lambda_k |e_k\rangle \langle e_k| | e_j \rangle = \sum_{k=1}^{\dim \mathcal{H}} \lambda_k \langle e_j | e_k \rangle \langle e_k | e_j \rangle \\ &= \sum_{k=1}^{\dim \mathcal{H}} \lambda_k \delta_{jk} = \lambda_j. \end{aligned}$$

Moreover, the trace condition implies that

$$\begin{aligned} 1 = \text{tr}(\rho) &= \sum_{j=1}^{\dim \mathcal{H}} \langle e_j | \rho | e_j \rangle = \sum_{j=1}^{\dim \mathcal{H}} \langle e_j | \sum_{k=1}^{\dim \mathcal{H}} \lambda_k |e_k\rangle \langle e_k| | e_j \rangle \\ &= \sum_{j=1}^{\dim \mathcal{H}} \sum_{k=1}^{\dim \mathcal{H}} \lambda_k \langle e_j | e_k \rangle \langle e_k | e_j \rangle = \sum_{j=1}^{\dim \mathcal{H}} \sum_{k=1}^{\dim \mathcal{H}} \lambda_k \delta_{jk} = \sum_{j=1}^{\dim \mathcal{H}} \lambda_j. \end{aligned}$$

Therefore, $\{\lambda_j\}_{j=1}^{\dim \mathcal{H}}$ is a probability distribution, and since the $|e_j\rangle$ are normalised, Equation (5.3) is a density operator corresponding to a probabilistic mixture of states. \square

Furthermore, taking probabilistic mixtures of density operators again leads to a density operator:

PROPOSITION 5.8. *If $\rho \in \mathcal{L}(\mathcal{H})$ and $\sigma \in \mathcal{L}(\mathcal{H})$ are density operators, then $p\rho + (1-p)\sigma$ is also a density operator for all $p \in [0, 1]$.*

PROOF. According to the theorem above, we need to prove positive semidefiniteness and the trace property. For every $|\psi\rangle \in \mathcal{H}$, we have

$$\langle \psi | [p\rho + (1-p)\sigma] | \psi \rangle = p\langle \psi | \rho | \psi \rangle + (1-p)\langle \psi | \sigma | \psi \rangle \geq 0,$$

since $\rho \geq 0$ and $\sigma \geq 0$. Furthermore,

$$\text{tr}[p\rho + (1-p)\sigma] = p\text{tr}(\rho) + (1-p)\text{tr}(\sigma) = p + (1-p) = 1$$

which proves the proposition. \square

In the following, therefore, we will use the term *density operator* or *density matrix* for any positive semidefinite operator with trace one. The terminology *mixed states* is used when ρ is a probabilistic mixture of more than one states.

DEFINITION 5.9. If a density operator $\rho \in \mathcal{L}(\mathcal{H})$ can be written as $\rho = |\psi\rangle\langle\psi|$ for some $|\psi\rangle \in \mathcal{H}$, then ρ is a *pure state*. Otherwise, it is called a *mixed state*. \diamond

While given a generic density operator it may seem at first difficult to decide whether it is pure or not, the following proposition provides a useful characterisation of pure states.

PROPOSITION 5.10. For every density operator ρ it holds that $\text{tr}(\rho^2) \leq 1$ with equality if and only if ρ is a pure state.

5.3 POSTULATES REVISITED

Since probabilistic mixtures of quantum states are just as valid as pure states, we revisit the postulates of quantum theory now taking the fundamental description of a quantum system to be a density operator.

POSTULATE (STATES). A (quantum mechanical) system is associated with a complex Hilbert space \mathcal{H} . The *state* of the system is represented by a positive semidefinite operator on this Hilbert space with unit trace, that is, by $\rho \in \mathcal{L}(\mathcal{H})$ such that $\rho \geq 0$ and $\text{tr}(\rho) = 1$. \diamond

POSTULATE (MEASUREMENTS). Consider a physical system with an associated Hilbert space \mathcal{H} . A measurement with n possible outcomes corresponds to a collection of n operators (one representing each outcome) on \mathcal{H} , that is, a set $\{M_j\}_{j=1}^n$ with $M_j \in \mathcal{L}(\mathcal{H})$.

If the state of the system before the measurement is $\rho \in \mathcal{L}(\mathcal{H})$, then the probability of obtaining an outcome j is given by

$$p(j) = \text{tr}(M_j^\dagger M_j \rho).$$

After observing outcome j , the updated state of the system is given by

$$\sigma_j = \frac{M_j \rho M_j^\dagger}{\text{tr}(M_j^\dagger M_j \rho)} = \frac{M_j \rho M_j^\dagger}{p(j)}.$$

The measurement operators must satisfy the *completeness relation*,

$$\sum_{j=1}^n M_j^\dagger M_j = \mathbf{1}.$$

\diamond

POSTULATE (DYNAMICS). The time evolution of a *closed* quantum system associated to a Hilbert space \mathcal{H} is described by a unitary operator $U \in \mathcal{L}(\mathcal{H})$. That is, the state $\rho(t)$ of a system at time t is related to the state $\rho(t')$ of the system at time t' by a unitary operator U that only depends on t and t' ,

$$\rho(t') = U\rho(t)U^\dagger. \quad \diamond$$

POSTULATE (COMPOSITE SYSTEMS). Consider a physical system associated to the Hilbert space \mathcal{H} , and another physical system associated to the Hilbert space \mathcal{K} . Then, the joint system is described by the *tensor product* of the two Hilbert spaces, $\mathcal{H} \otimes \mathcal{K}$. The joint state of $\rho \in \mathcal{L}(\mathcal{H})$ and $\sigma \in \mathcal{L}(\mathcal{K})$ is described by $\rho \otimes \sigma \in \mathcal{L}(\mathcal{H} \otimes \mathcal{K})$. \diamond

REMARK 5.11. Note that in the measurement postulate, the numbers $p(j)$ indeed form a probability distribution. Indeed, if $\rho = \sum_k q_k |\psi_k\rangle\langle\psi_k|$ then

$$\begin{aligned} p(j) &= \text{tr} \left(M_j^\dagger M_j \sum_k q_k |\psi_k\rangle\langle\psi_k| \right) = \sum_k q_k \text{tr}(M_j^\dagger M_j |\psi_k\rangle\langle\psi_k|) \\ &= \sum_k q_k \langle\psi_k| M_j^\dagger M_j |\psi_k\rangle = \sum_k q_k \|M_j |\psi_k\rangle\|^2 \geq 0. \end{aligned}$$

and due to the completeness relation, we have

$$\sum_{j=1}^n p(j) = \sum_{j=1}^n \text{tr}(M_j^\dagger M_j \rho) = \text{tr} \left(\sum_{j=1}^n M_j^\dagger M_j \rho \right) = \text{tr}(\mathbb{1}\rho) = \text{tr}(\rho) = 1.$$

Furthermore, the post-measurement states σ_j are indeed valid density operators, since for every $|\varphi\rangle \in \mathcal{H}$, we have

$$\begin{aligned} \langle\varphi|\sigma_j|\varphi\rangle &= \frac{\langle\varphi|M_j\rho M_j^\dagger|\varphi\rangle}{p(j)} = \frac{1}{p(j)} \langle\varphi|M_j \sum_k q_k |\psi_k\rangle\langle\psi_k|M_j^\dagger|\varphi\rangle \\ &= \frac{1}{p(j)} \sum_k q_k \langle\varphi|M_j |\psi_k\rangle\langle\psi_k|M_j^\dagger|\varphi\rangle = \frac{1}{p(j)} \sum_k q_k |\langle\varphi|M_j |\psi_k\rangle|^2 \geq 0 \end{aligned}$$

and

$$\text{tr}(\sigma_j) = \frac{\text{tr}(M_j\rho M_j^\dagger)}{\text{tr}(M_j^\dagger M_j \rho)} = 1. \quad \diamond$$

5.4 BLOCH BALL

In the case of qubits, generalising from pure states to mixed states generalises the notion of the Bloch sphere to the Bloch *ball*:

PROPOSITION 5.12. Consider a qubit density matrix, $\rho \in \mathcal{L}(\mathbb{C}^2)$ such that $\rho \geq 0$ and $\text{tr}(\rho) = 1$. Any such matrix can be written as

$$\rho = \frac{1}{2}(\mathbb{1} + \vec{r} \cdot \vec{\sigma}),$$

where $\vec{r} = (x, y, z) \in \mathbb{R}^3$ is a vector with $\|\vec{r}\| = \sqrt{x^2 + y^2 + z^2} \leq 1$, and $\vec{\sigma} = (X, Y, Z)$ is a vector containing the Pauli matrices, and therefore $\vec{r} \cdot \vec{\sigma} = xX + yY + zZ$. The vector \vec{r} is called the Bloch vector for the state ρ , and the set of all states corresponds to the unit ball in terms of the set of Bloch vectors.

A state is pure, $\rho = |\psi\rangle\langle\psi|$ if and only if $\|\vec{r}\| = 1$. If the state $|\psi\rangle$ is represented on the Bloch sphere by the angles (θ, ϕ) then the corresponding Bloch vector is

$$\vec{r} = (\sin \theta \cos \phi, \sin \theta \sin \phi, \cos \theta).$$

PROOF. See the problem class. \square

5.5 REDUCED DENSITY OPERATOR

Apart from providing a description of probabilistic mixtures of quantum states, the density operator formalism allows for the description of subsystems of some composite system. We saw that this can be problematic in the case of entangled states. The density operator formalism provides the proper description of the subsystems of even an entangled state. In order to be able to define these *reduced states* of composite systems, we need the concept of the *partial trace*.

DEFINITION 5.13. Let $A \in \mathcal{L}(\mathcal{H}_1)$ and $B \in \mathcal{L}(\mathcal{H}_2)$ be linear operators on the Hilbert spaces \mathcal{H}_1 and \mathcal{H}_2 , respectively. The *partial trace over system 1* of the tensor product operator $A \otimes B \in \mathcal{L}(\mathcal{H}_1 \otimes \mathcal{H}_2)$ is given by

$$\text{tr}_1(A \otimes B) = \text{tr}(A)B \in \mathcal{L}(\mathcal{H}_2).$$

The partial trace of a generic operator on $\mathcal{H}_1 \otimes \mathcal{H}_2$ is the linear extension of the above definition. That is, for an arbitrary operator $C = \sum_i A_i \otimes B_i$, the partial trace over system 1 is given by

$$\text{tr}_1(C) = \sum_i \text{tr}(A_i)B_i \in \mathcal{L}(\mathcal{H}_2).$$

The partial trace over system 2 is defined in an analogous way,

$$\text{tr}_2(C) = \sum_i \text{tr}(B_i)A_i \in \mathcal{L}(\mathcal{H}_1).$$

\diamond

REMARK 5.14. The partial trace is linear,

$$\text{tr}_1(\alpha A + \beta B) = \alpha \text{tr}_1(A) + \beta \text{tr}_1(B) \quad \forall A, B \in \mathcal{L}(\mathcal{H}_1 \otimes \mathcal{H}_2), \forall \alpha, \beta \in \mathbb{C},$$

it is cyclic (on the Hilbert space over which the trace is taken)

$$\mathrm{tr}_1[(A \otimes \mathbb{1})B] = \mathrm{tr}_1[B(A \otimes \mathbb{1})] \quad \forall A \in \mathcal{L}(\mathcal{H}_1), \forall B \in \mathcal{L}(\mathcal{H}_1 \otimes \mathcal{H}_2)$$

and it has the tower property:

$$\mathrm{tr}[\mathrm{tr}_1(A)] = \mathrm{tr}(A) \quad \forall A \in \mathcal{L}(\mathcal{H}_1 \otimes \mathcal{H}_2). \quad \diamond$$

Now we can define the reduced state of a density operator on a tensor product Hilbert space.

DEFINITION 5.15. Consider a density operator ρ on the tensor product Hilbert space $\mathcal{H}_1 \otimes \mathcal{H}_2$. The density operator describing the state of the system on the subsystem 1 (the *reduced state on system 1*) is given by

$$\rho_1 = \mathrm{tr}_2(\rho) \in \mathcal{L}(\mathcal{H}_1),$$

and the density operator describing the state of the system on the subsystem 2 (the *reduced state on system 2*) is given by

$$\rho_2 = \mathrm{tr}_1(\rho) \in \mathcal{L}(\mathcal{H}_2). \quad \diamond$$

To make sure that this definition is appropriate, we need to ensure that the reduced state is indeed a quantum state.

PROPOSITION 5.16. *If $\rho \in \mathcal{L}(\mathcal{H}_1 \otimes \mathcal{H}_2)$ is a density operator, then the reduced states $\rho_1 \in \mathcal{L}(\mathcal{H}_1)$ and $\rho_2 \in \mathcal{L}(\mathcal{H}_2)$ are also density operators.*

PROOF. Consider the reduced state $\rho_1 = \mathrm{tr}_2(\rho)$. Given an orthonormal basis $\{|e_i\rangle\}_{i=1}^{\dim \mathcal{H}_2}$ on \mathcal{H}_2 , the partial trace can be written as

$$\mathrm{tr}_1(\rho) = \sum_{i=1}^{\dim \mathcal{H}_2} E_i \rho E_i^\dagger,$$

where $E_i = \mathbb{1} \otimes \langle e_i|$. Then, for any $|\psi\rangle \in \mathcal{H}_2$, we have

$$\langle \psi | \rho_1 | \psi \rangle = \sum_{i=1}^{\dim \mathcal{H}_2} \langle \psi | E_i \rho E_i^\dagger | \psi \rangle = \sum_{i=1}^{\dim \mathcal{H}_2} \langle E_i^\dagger \psi | \rho | E_i^\dagger \psi \rangle \geq 0$$

since $\rho \geq 0$, and therefore $\rho_1 \geq 0$. For the trace property, note that

$$\begin{aligned} \mathrm{tr}(\rho_1) &= \sum_{i=1}^{\dim \mathcal{H}_2} \mathrm{tr}(E_i \rho E_i^\dagger) = \sum_{i=1}^{\dim \mathcal{H}_2} \mathrm{tr}(\rho E_i^\dagger E_i) \\ &= \sum_{i=1}^{\dim \mathcal{H}_2} \mathrm{tr}[\rho(\mathbb{1} \otimes |e_i\rangle\langle e_i|)] = \mathrm{tr} \left[\rho \left(\mathbb{1} \otimes \sum_{i=1}^{\dim \mathcal{H}_2} |e_i\rangle\langle e_i| \right) \right] \\ &= \mathrm{tr}[\rho(\mathbb{1} \otimes \mathbb{1})] = \mathrm{tr}(\rho) = 1, \end{aligned}$$

and therefore ρ_1 is a density operator. An analogous proof holds for ρ_2 . \square

It is clear that the reduced density operator gives the correct state of the subsystems of a product state:

EXAMPLE 5.17. If $\rho = \sigma_1 \otimes \sigma_2$, then the reduced states are

$$\rho_1 = \text{tr}_2(\sigma_1 \otimes \sigma_2) = \text{tr}(\sigma_2)\sigma_1 = \sigma_1,$$

$$\rho_2 = \text{tr}_1(\sigma_1 \otimes \sigma_2) = \text{tr}(\sigma_1)\sigma_2 = \sigma_2. \quad \diamond$$

To justify the use of the partial trace for describing the subsystems of arbitrary (even entangled) states, consider a composite system $\rho \in \mathcal{L}(\mathcal{H}_1 \otimes \mathcal{H}_2)$, and a local measurement described by the POVM $\{M_j\}_{j=1}^n$ on \mathcal{H}_1 . The outcome probabilities (according to the postulates of quantum theory) are given by

$$p(j) = \text{tr}[(M_j \otimes \mathbb{1})\rho].$$

Therefore, if we want to describe the subsystem on \mathcal{H}_1 compatible with these statistics, we need a linear map, $f : \mathcal{L}(\mathcal{H}_1 \otimes \mathcal{H}_2) \rightarrow \mathcal{L}(\mathcal{H}_1)$ such that

$$p(j) = \text{tr}[M_j f(\rho)]$$

for all density operators ρ on $\mathcal{H}_1 \otimes \mathcal{H}_2$. The following theorem proves that the partial trace is the unique option for this linear map, that is, $f(\rho) = \text{tr}_2(\rho)$.



THEOREM 5.18. Consider a linear map $f : \mathcal{L}(\mathcal{H}_1 \otimes \mathcal{H}_2) \rightarrow \mathcal{L}(\mathcal{H}_1)$ such that for all density operators ρ on $\mathcal{H}_1 \otimes \mathcal{H}_2$ and for all $M \in \mathcal{L}(\mathcal{H}_1)$ it holds that

$$\text{tr}[(M \otimes \mathbb{1})\rho] = \text{tr}[M f(\rho)].$$

Then, we have that $f(\rho) = \text{tr}_2(\rho)$.

The reduced state allows us to characterise the subsystems of entangled states. Take, for example, the Bell state $|\phi^+\rangle$:

EXAMPLE 5.19. Consider the state $|\phi^+\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle) \in \mathbb{C}^2 \otimes \mathbb{C}^2$, and the corresponding density operator,

$$|\phi^+\rangle\langle\phi^+| = \frac{1}{2}(|00\rangle + |11\rangle)(\langle 00| + \langle 11|).$$

The reduced state on the first qubit is given by

$$\begin{aligned} \text{tr}_2(|\phi^+\rangle\langle\phi^+|) &= \frac{1}{2} \text{tr}_2(|00\rangle\langle 00| + |00\rangle\langle 11| + |11\rangle\langle 00| + |11\rangle\langle 11|) \\ &= \frac{1}{2}(\langle 0|0\rangle |0\rangle\langle 0| + \langle 1|0\rangle |0\rangle\langle 1| + \langle 0|1\rangle |1\rangle\langle 0| + \langle 1|1\rangle |1\rangle\langle 1|) \\ &= \frac{1}{2}(|0\rangle\langle 0| + |1\rangle\langle 1|) = \frac{\mathbb{1}}{2}. \end{aligned}$$

This is the *maximally mixed state* from Example 5.6. This is somewhat paradoxical, since the state of the joint system contains a lot of information (2 bits, if considered as one of the elements of an orthonormal basis on $\mathbb{C}^2 \otimes \mathbb{C}^2$), but the subsystems do not contain any useful information [a similar computation shows that $\text{tr}_1(|\phi^+\rangle\langle\phi^+|) = \frac{\mathbb{1}}{2}$]. This is one of the peculiar consequences of entanglement. \diamond

Note that the reduced state and density operators in general become essential when we want to describe the time evolution of a system that is not closed. Consider Example 4.25 again, and focus on the first qubit, whose initial state is $|0\rangle$. Then, it undergoes a unitary time evolution described by H , and the updated state is $|+\rangle$. After this, the qubit interacts with its *environment*, another qubit in the state $|0\rangle$, and *the joint state* $|+\rangle \otimes |0\rangle$ undergoes a unitary evolution described by the CNOT gate, and the updated state is $|\phi^+\rangle$. The state of the first qubit is now given by the density operator $\frac{1}{2}$.

The time evolution of the first qubit—in the density operator language—is $|0\rangle\langle 0| \rightarrow \frac{1}{2}$ in the above procedure. One can show that this time evolution is in fact *not* unitary, which is a consequence of the fact that the qubit system is not closed (it interacts with another qubit via the CNOT gate). It is also the consequence of the system being not closed that the final state cannot be described by a pure state, and it becomes mixed.

5.6 ENTANGLEMENT AND NO-SIGNALLING

The notion of entanglement can be extended to mixed states. Recall that we call a pure state a product state if it is a tensor product of two states, $|\psi\rangle \otimes |\varphi\rangle$. Otherwise, it is entangled. The density operator of a product state is given by a tensor product of density operators, $|\psi\rangle\langle\psi| \otimes |\varphi\rangle\langle\varphi|$. We will call such a density operator a product state as well. Note that in the density operator formalism, we can take probabilistic mixtures of these product states. Such states are called *separable*, and these are the proper generalisation of product states. This is because a probabilistic (classical) mixture of product states cannot be considered entangled, as entanglement measures the purely quantum correlation between subsystems. The formal definition is as follows:

DEFINITION 5.20. Let $\mathcal{H}_1, \dots, \mathcal{H}_n$ be Hilbert spaces, and $\rho \in \mathcal{L}(\mathcal{H}_1 \otimes \dots \otimes \mathcal{H}_n)$ a density operator. If there exist density operators $\rho_j^i \in \mathcal{L}(\mathcal{H}_j)$ and a probability distribution $\{p_i\}$ such that

$$\rho = \sum_i p_i \rho_1^i \otimes \dots \otimes \rho_n^i \quad (5.4)$$

then ρ is called *separable*. Otherwise it is *entangled*. \diamond

That is, the generalisation of product states to density operators is separable states (i.e. probabilistic mixtures of product states).

REMARK 5.21. Note that the states ρ_j^i in Equation (5.4) can be taken to be pure, without loss of generality (if some of them are mixtures of pure states, the probabilities in the mixtures can be absorbed in the p_i). \diamond

EXAMPLE 5.22. Any pure entangled state is still entangled according to Definition 5.20. In particular, the Bell state $|\phi^+\rangle\langle\phi^+|$ cannot be written as a convex combination (probabilistic mixture) of product states. \diamond

For general mixed states, deciding whether a state is entangled or not is a very difficult (NP-hard) problem. While there exist methods that detect the entanglement of specific states, these are beyond the scope of this module. As such, in the following we will look at a few fundamental aspects of quantum theory that can be described using entanglement and mixed states.

We saw at the end of Chapter 4 that locally measuring a bipartite quantum state immediately changes the state of the system through the post-measurement state. E.g. the state $|\phi^+\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$ changes to $|00\rangle$ if Alice measures the first qubit in the computational basis and observes outcome ‘0’. At first sight, this seems to imply the possibility of faster-than-light communication: all Alice needs to do is measure her qubit, and the state (and therefore the qubit of Bob) immediately changes. Note, however, that Bob doesn’t know Alice’s measurement outcome (and hence the updated quantum state) unless Alice communicates this to him, and this communication cannot be done faster than light. Thus, the theory of relativity seems to be saved. This impossibility of immediate communication can be made rigorous using density operators:

THEOREM 5.23 (NO-SIGNALLING PRINCIPLE). *Alice cannot communicate instantaneously to Bob by measuring her part of an entangled quantum state.*

REMARK 5.24. The above is sometimes called the no-communication theorem. \diamond

PROOF. Imagine that Alice and Bob share a bipartite quantum state, $\rho \in \mathcal{L}(\mathcal{H}_A \otimes \mathcal{H}_B)$, and Alice performs an n -outcome measurement described by the measurement operators $\{A_a\}_{a=1}^n$ on \mathcal{H}_A such that $\sum_{a=1}^n A_a^\dagger A_a = \mathbb{1}$. The measurement operators on the tensor product Hilbert space are therefore given by $\{A_a \otimes \mathbb{1}\}_{a=1}^n$. According to the measurement postulate, the probability of Alice observing outcome a is $p(a) = \text{tr}[(A_a^\dagger A_a \otimes \mathbb{1})\rho]$, and the post-measurement state after observing outcome a is

$$\sigma_a = \frac{1}{p(a)}(A_a \otimes \mathbb{1})\rho(A_a^\dagger \otimes \mathbb{1}).$$

Let us compute the post-measurement state of Bob, after Alice has performed a measurement. Note that Bob doesn’t know the outcome, and therefore the post-measurement state from his perspective is a probabilistic mixture of the post-measurement states conditioned on the different outcomes. Furthermore, we need to compute the reduced state on Bob’s side (and since the partial trace is linear, it doesn’t matter in which order we take the probabilistic mixture and the partial trace). Bob’s state is therefore given by

$$\begin{aligned} \rho_B &= \text{tr}_A \left[\sum_{a=1}^n p(a) \sigma_a \right] = \text{tr}_A \left[\sum_{a=1}^n (A_a \otimes \mathbb{1}) \rho (A_a^\dagger \otimes \mathbb{1}) \right] \\ &= \text{tr}_A \left[\sum_{a=1}^n \rho (A_a^\dagger A_a \otimes \mathbb{1}) \right] = \text{tr}_A [\rho (\mathbb{1} \otimes \mathbb{1})] = \text{tr}_A (\rho), \end{aligned}$$

where we used the form of the post-measurement state, the cyclicity of the partial trace, and the completeness relation of the measurement operators. Therefore, Alice's measurement doesn't change the reduced state of Bob, meaning that it is impossible for Alice to communicate anything to Bob by performing a measurement on her part of the system (unless she communicates the measurement outcome). \square

5.7 PURIFICATION OF STATES

The sections above show that density operators provide a very useful characterisation of quantum systems in various physical scenarios (such as probabilistic mixtures or sub-systems of composite systems). In this section, we will see that in fact every mixed state can be seen as a subsystem of some larger (potentially entangled) pure state. As such, one might think of pure states as the fundamental description of quantum systems, and mixed states as a convenient description of subsystems or systems of which we have limited information (and therefore we resort to describing it as a probabilistic mixture of pure states).

THEOREM 5.25. *For every mixed state $\rho \in \mathcal{L}(\mathcal{H})$, there exist a Hilbert space \mathcal{K} , and a pure state $|\psi\rangle \in \mathcal{H} \otimes \mathcal{K}$ such that $\rho = \text{tr}_{\mathcal{K}}(|\psi\rangle\langle\psi|)$. Such a vector $|\psi\rangle$ is called a purification of ρ .*

PROOF. Since ρ is Hermitian, by the spectral theorem we can write it as

$$\rho = \sum_{j=1}^{\dim \mathcal{H}} \lambda_j |\varphi_j\rangle\langle\varphi_j|,$$

where λ_j is a probability distribution (since $\rho \geq 0$ and $\text{tr}(\rho) = 1$) and $\{|\varphi_j\rangle\}_{j=1}^{\dim \mathcal{H}}$ is an orthonormal basis of \mathcal{H} . Consider the state

$$|\psi\rangle = \sum_{j=1}^{\dim \mathcal{H}} \sqrt{\lambda_j} |\varphi_j\rangle \otimes |j\rangle,$$

where $\{|j\rangle\}_{j=1}^{\dim \mathcal{H}}$ is an orthonormal set (e.g. the computational basis) on some Hilbert space \mathcal{K} (from this, we see that we can always choose $\mathcal{K} \simeq \mathcal{H}$).

Then, we have

$$\begin{aligned} \text{tr}_{\mathcal{K}}(|\psi\rangle\langle\psi|) &= \text{tr}_{\mathcal{K}} \left[\left(\sum_{j=1}^{\dim \mathcal{H}} \sqrt{\lambda_j} |\varphi_j\rangle \otimes |j\rangle \right) \left(\sum_{k=1}^{\dim \mathcal{H}} \sqrt{\lambda_k} \langle\varphi_k| \otimes \langle k| \right) \right] \\ &= \sum_{j=1}^{\dim \mathcal{H}} \sum_{k=1}^{\dim \mathcal{H}} \sqrt{\lambda_j} \sqrt{\lambda_k} \text{tr}_{\mathcal{K}} (|\varphi_j\rangle\langle\varphi_k| \otimes |j\rangle\langle k|) \\ &= \sum_{j=1}^{\dim \mathcal{H}} \sum_{k=1}^{\dim \mathcal{H}} \sqrt{\lambda_j} \sqrt{\lambda_k} |\varphi_j\rangle\langle\varphi_k| \text{tr}(|j\rangle\langle k|) \\ &= \sum_{j=1}^{\dim \mathcal{H}} \lambda_j |\varphi_j\rangle\langle\varphi_j| = \rho, \end{aligned}$$

where we used that $\text{tr}(|j\rangle\langle k|) = \langle k|j\rangle = \delta_{jk}$. \square

EXAMPLE 5.26. A purification of the mixed single-qubit state

$$\rho = \frac{3}{4} |+\rangle\langle+| + \frac{1}{4} |-\rangle\langle-|$$

is given by the pure two-qubit state

$$|\psi\rangle = \frac{\sqrt{3}}{2} |+\rangle \otimes |0\rangle + \frac{1}{2} |-\rangle \otimes |1\rangle. \quad \diamond$$

In a similar vein, every POVM can be viewed as a projective measurement (PVM) on a bigger system (the analogue of a purification in this case is called *Naimark dilation*). That is, projective measurements can be thought of as the fundamental description of quantum measurements, and POVMs as describing a projective measurement on a subsystem (and this is how non-projective POVMs are implemented in practice). In quantum information theory, however, we usually allow for the most general description of states and measurements in terms of density operators and POVMs, and the purification of the state and the dilation of the measurement to a projective measurement is secondary.

6 Quantum circuits

In this chapter, we start formulating quantum information processing protocols as circuits, in preparation for the description of quantum algorithms in the subsequent chapter. Quantum circuits take a multi-qubit state as an input, and apply various gates (single- and multi-qubit unitary operators) on this initial state. While quantum circuits can be defined for density operators, for simplicity (and in order to model ideal implementations), we will mostly consider pure states in the following, and therefore we will work in the Hilbert space vector formalism instead of the density operator formalism.

6.1 NO-CLONING THEOREM

As a warm-up, before introducing the circuit notation, we prove a fundamental theorem about unitary operators called the *no-cloning theorem*. This peculiar property of quantum theory makes quantum information processing fundamentally different from classical information processing. No-cloning is behind some of the fundamental advantages of quantum information processing, but it also causes some difficulties. In simple terms, no-cloning means that—in stark contrast to classical information—quantum information cannot be perfectly copied.

THEOREM 6.1. *Let \mathcal{H} be a Hilbert space. There exists no unitary operator U on $\mathcal{H} \otimes \mathcal{H}$ and state $|w\rangle \in \mathcal{H}$ such that¹¹*

$$U(|v\rangle \otimes |w\rangle) \sim |v\rangle \otimes |v\rangle$$

for all states $|v\rangle \in \mathcal{H}$.

PROOF. We prove the theorem by contradiction. Consider two arbitrary states $|v_1\rangle, |v_2\rangle \in \mathcal{H}$, and assume that there exists a unitary operator U and a state $|w\rangle \in \mathcal{H}$ such that

$$U(|v_i\rangle \otimes |w\rangle) \sim |v_i\rangle \otimes |v_i\rangle$$

for $i = 1, 2$. Then, we have

$$\begin{aligned} \langle v_1 | v_2 \rangle &= \langle v_1 | v_2 \rangle \langle w | w \rangle = (\langle v_1 | \otimes \langle w |) (|v_2\rangle \otimes |w\rangle) \\ &= (\langle v_1 | \otimes \langle w |) U^\dagger U (|v_2\rangle \otimes |w\rangle) \\ &= e^{i\phi} (\langle v_1 | \otimes \langle v_1 |) (|v_2\rangle \otimes |v_2\rangle) \\ &= e^{i\phi} \langle v_1 | v_2 \rangle^2 \end{aligned}$$

¹¹ As before, \sim denotes equivalence of states, that is, equality up to a global phase.

for some phase $\phi \in \mathbb{R}$. This implies that either $\langle v_1 | v_2 \rangle = 0$ or $\langle v_1 | v_2 \rangle = e^{-i\phi}$, which is not true for *arbitrary* states. \square



REMARK 6.2. As the theorem above shows, it is impossible to construct an apparatus that perfectly clones arbitrary states. It is, however, possible to clone arbitrary states imperfectly using a variety of ‘machines’.¹² \diamond

¹² V. Bužek and M. Hillery.
“Quantum Copying: Beyond the
No-Cloning Theorem”.
Phys. Rev. A 54 (3),
pp. 1844–1852, 1996.

Note that states that are orthogonal or equivalent to each other *can* be copied. This is equivalent to the ability to clone classical information (by e.g. associating the two states $|0\rangle$ and $|1\rangle$ to a classical bit). In this sense, no-cloning is what’s behind the security of the BB84 protocol. Since the states Alice sends in the protocol $\{|0\rangle, |1\rangle, |+\rangle, |-\rangle\}$ are not all orthogonal or equivalent to each other, Eve cannot copy them and therefore she cannot remain unnoticed if she tries to learn something about these states.

6.2 QUANTUM GATES AND CIRCUIT DIAGRAMS

We now introduce a graphical representation of quantum gates, and using these, we draw circuit diagrams that can be used to describe quantum information processing protocols and quantum algorithms. Quantum gates are just unitary operators, and we have already seen a couple of quantum gates in the previous chapters. In particular, we have introduced in Section 2.3 the single-qubit gates X, Y, Z and H —the Pauli gates and the Hadamard gate—and in Section 4.4 the two-qubit gates C_X and SWAP—the controlled NOT gate and the SWAP gate.

It is useful to introduce a graphical notation for products and tensor products of quantum gates. In this graphical notation, we represent a single qubit gate U as

$$U |v\rangle = |w\rangle \quad \longleftrightarrow \quad |v\rangle \text{ --- } \boxed{U} \text{ --- } |w\rangle = U |v\rangle.$$

The identity is simply represented by a ‘wire’

$$\mathbb{1} |v\rangle = |v\rangle \quad \longleftrightarrow \quad |v\rangle \text{ ————— } |v\rangle.$$

Multiple-qubit gates are represented using multiple input and output wires stacked on top of each other, e.g., for a three-qubit gate U

$$U |v\rangle = |w\rangle \quad \longleftrightarrow \quad |v\rangle \left\{ \begin{array}{c} \text{---} \\ \text{---} \\ \text{---} \end{array} \right\} \boxed{U} \left\{ \begin{array}{c} \text{---} \\ \text{---} \\ \text{---} \end{array} \right\} |w\rangle = U |v\rangle.$$

Sometimes when drawing a circuit for arbitrary numbers of qubits, it is useful to bundle multiple wires. Bundles of wires are lined up and indicated with a slash on the wire. For example, the last circuit can also be represented by

$$|v\rangle \text{ --- } \text{ }^3\text{---} \boxed{U} \text{ --- } |w\rangle = U |v\rangle.$$

Combinations of multiple gates using compositions and tensor products, can also be represented in this graphical notation, as shown, e.g., in Figure 6.1.

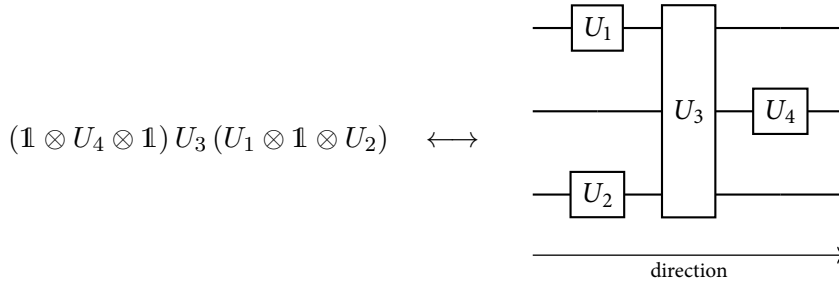


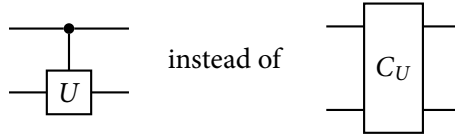
FIGURE 6.1. A three-qubit circuit consisting of four quantum gates. The direction in which the circuit has to be read is indicated.

Note that the order of gates in the circuit is from the left to the right, while it is from the right to the left in the corresponding operator expression. To emphasise this, we have included in all the preceding circuits input and output kets. These will be omitted in most of the following circuit diagrams, where we only display the part of the diagram corresponding to the operator.

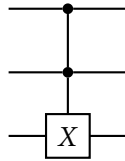
Controlled gates (with the control on the first qubit), in operator representation given by

$$C_U = |0\rangle\langle 0| \otimes \mathbb{1} + |1\rangle\langle 1| \otimes U$$

are represented diagrammatically as



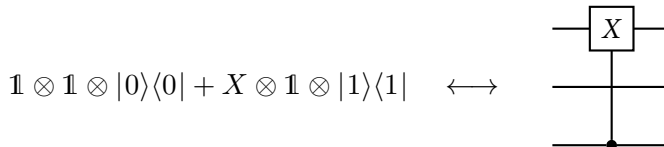
The filled dot on the top wire indicates that the first qubit is the control qubit. This notation is very flexible and allows the representation of gates with controls and targets on arbitrary wires. For example, the *Toffoli gate* is represented by



which translates to the operator representation

$$(|00\rangle\langle 00| + |01\rangle\langle 01| + |10\rangle\langle 10|) \otimes \mathbb{1} + |11\rangle\langle 11| \otimes X.$$

Furthermore, a CNOT gate with control on the third qubit and target on the first qubit is given by



CAUTION. The notions of control and target in controlled gates is not as clear-cut as it might seem at first sight. Compare, for example, the application of the CNOT gate C_X to the computational basis states and to the Hadamard basis states:

$$\left. \begin{array}{l} |00\rangle \\ |01\rangle \\ |10\rangle \\ |11\rangle \end{array} \right\} \xrightarrow{C_X} \left\{ \begin{array}{l} |00\rangle \\ |01\rangle \\ |11\rangle \\ |10\rangle \end{array} \right. \quad \text{compared to} \quad \left. \begin{array}{l} |++\rangle \\ |+-\rangle \\ |-+\rangle \\ |--\rangle \end{array} \right\} \xrightarrow{C_X} \left\{ \begin{array}{l} |++\rangle \\ |--\rangle \\ |-+\rangle \\ |+-\rangle \end{array} \right.$$

The action in the computational basis supports the naïve interpretation of the control and target bits. The target bit is changed depending on the value of the control bit, while the control bit is left unchanged. However, when applying the C_X gate to the Hadamard basis states, the roles of the control and target bits apparently reverse. Related to this is the identity

$$(H \otimes H) C_X (H \otimes H) = \mathbb{1} \otimes |0\rangle\langle 0| + X \otimes |1\rangle\langle 1|,$$

represented graphically in Figure 6.2.

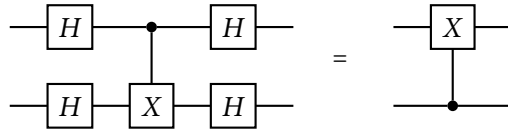
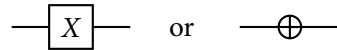
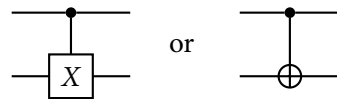


FIGURE 6.2. The control and target bits of a controlled NOT gate can be swapped using Hadamard gates.

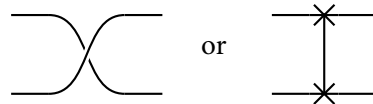
It is important to be aware that some frequently occurring gates have alternative representations. The Pauli X -gate has the two alternative representations



and therefore the CNOT gate is represented as



The SWAP gate has the two alternative representations



Circuit diagrams can also accommodate measurements. Measurements of one or multiple qubits in the computational basis are represented by a meter:



The input of the measurement is a quantum system (represented by a wire), while the output is a classical label (represented by a double *classical wire*, carrying the measurement result).

REMARK 6.3. Usually when we consider a quantum circuit, we use the input state

$$|0\rangle^{\otimes n} := |0\rangle \otimes \cdots \otimes |0\rangle,$$

that is, the n -fold tensor product of a fixed qubit state that we choose to be the computational basis $|0\rangle$ state. The idea is that there is a fixed qubit state that we can prepare with high accuracy, and every other state is prepared from this one by the action of some unitaries, described by the quantum circuit. \diamond

6.3 QUANTUM TELEPORTATION

We formulate our first information processing protocol as a quantum circuit. Quantum teleportation is a protocol that allows a party (usually called Alice) to “teleport” an *arbitrary* qubit state to another party (usually called Bob), using only two bits of classical communication and a pre-shared entangled Bell state. The protocol works as follows (also see the circuit diagram in Figure 6.3).

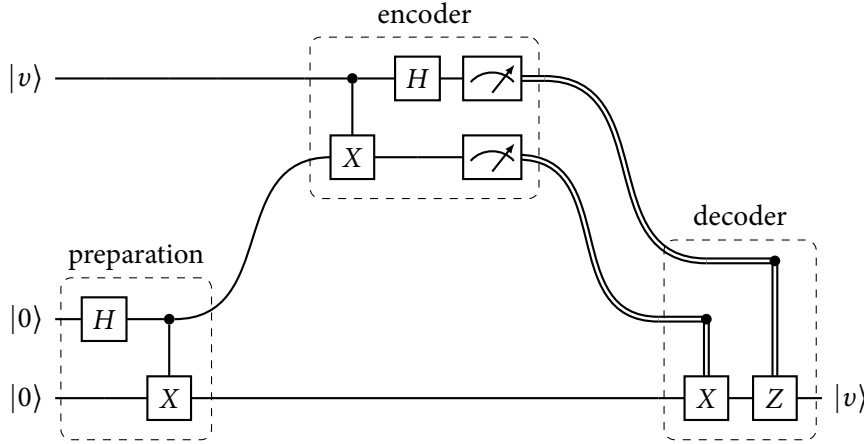


FIGURE 6.3. Circuit for quantum teleportation, where Alice is in the “encoder” box and Bob is in the “decoder” box. After a Bell pair $|\Phi^+\rangle$ is prepared, Alice measures the first qubit together with the qubit she wishes to teleport in the Bell basis. Bob then applies the X and Z gate on his qubit conditionally on Alice’s measurement outcomes.

PROTOCOL.

1. A Bell state $|\Phi^+\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$ is prepared. This is accomplished by applying $C_X(H \otimes \mathbb{1})$ to $|00\rangle$.
2. The first qubit of $|\Phi^+\rangle$ is sent to Alice, while the second qubit is sent to Bob. In addition, Alice holds the arbitrary qubit state $|v\rangle = \alpha|0\rangle + \beta|1\rangle$. The joint system of $|v\rangle$ and $|\Phi^+\rangle$ is therefore

$$\begin{aligned} |v\rangle \otimes |\Phi^+\rangle &= (\alpha|0\rangle + \beta|1\rangle) \otimes \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle) \\ &= \frac{1}{\sqrt{2}}(\alpha(|000\rangle + |011\rangle) + \beta(|100\rangle + |111\rangle)). \end{aligned}$$

Re-writing this state we obtain

$$\begin{aligned}
 |v\rangle \otimes |\Phi^+\rangle &= \frac{1}{2\sqrt{2}} \left(\alpha(|000\rangle + |110\rangle) + \beta(|001\rangle + |111\rangle) \right. \\
 &\quad + \alpha(|011\rangle + |101\rangle) + \beta(|010\rangle + |100\rangle) \\
 &\quad + \alpha(|000\rangle - |110\rangle) - \beta(|001\rangle - |111\rangle) \\
 &\quad \left. + \alpha(|011\rangle - |101\rangle) - \beta(|010\rangle - |100\rangle) \right) \\
 &= \frac{1}{2} \left(|\Phi^+\rangle \otimes (\alpha|0\rangle + \beta|1\rangle) + |\Psi^+\rangle \otimes (\alpha|1\rangle + \beta|0\rangle) \right. \\
 &\quad \left. + |\Phi^-\rangle \otimes (\alpha|0\rangle - \beta|1\rangle) + |\Psi^-\rangle \otimes (\alpha|1\rangle - \beta|0\rangle) \right).
 \end{aligned}$$

3. Alice measures her two qubits (i.e. the single-qubit system in the state she wants to ‘teleport’ and the qubit of the $|\phi^+\rangle$ state that was sent to her) in the Bell basis. She accomplishes this by first applying $(H \otimes \mathbb{1})C_X$ to her two qubits, that is, the three-qubit system changes to

$$\begin{aligned}
 (H \otimes \mathbb{1} \otimes \mathbb{1}) (C_X \otimes \mathbb{1})(|v\rangle \otimes |\Phi^+\rangle) \\
 = \frac{1}{2} \left(|00\rangle \otimes (\alpha|0\rangle + \beta|1\rangle) + |01\rangle \otimes (\alpha|1\rangle + \beta|0\rangle) \right. \\
 \left. + |10\rangle \otimes (\alpha|0\rangle - \beta|1\rangle) + |11\rangle \otimes (\alpha|1\rangle - \beta|0\rangle) \right).
 \end{aligned}$$

Then, Alice measures her qubits in the computational basis. Depending on the outcome $(a, b) \in \{0, 1\}^2$ of Alice’s measurement, Bob’s qubit will be in the state

$$\begin{cases} \alpha|0\rangle + \beta|1\rangle & \text{if } (a, b) = (0, 0), \\ \alpha|1\rangle + \beta|0\rangle & \text{if } (a, b) = (0, 1), \\ \alpha|0\rangle - \beta|1\rangle & \text{if } (a, b) = (1, 0), \\ \alpha|1\rangle - \beta|0\rangle & \text{if } (a, b) = (1, 1). \end{cases}$$

4. Alice sends the two classical bits encoding the outcome of the measurements to Bob.
5. Depending on Alice’s two-bit message (a, b) , Bob applies $Z^a X^b$ to his qubit (for any operator M , $M^1 = M$ and $M^0 = \mathbb{1}$). After this, his qubit state changes to

$$\left. \begin{array}{l} \alpha|0\rangle + \beta|1\rangle \\ \alpha|1\rangle + \beta|0\rangle \\ \alpha|0\rangle - \beta|1\rangle \\ \alpha|1\rangle - \beta|0\rangle \end{array} \right\} \xrightarrow{Z^a X^b} \alpha|0\rangle + \beta|1\rangle = |v\rangle.$$

Note that in this protocol we do not violate the no-cloning theorem: at no point is a state ‘copied’, since after the protocol Alice holds one of the computational basis states $\{|00\rangle, |01\rangle, |10\rangle, |11\rangle\}$.

6.4 DECOMPOSITION OF UNITARY OPERATORS

In this section, we look at how to decompose complicated multi-qubit unitaries into a sequence of single-qubit and two-qubit unitaries. We will start by some useful *rotation* unitaries:

DEFINITION 6.4. For $U = X, Y, Z$, define the *rotation operators*

$$R_U(\theta) := \cos \frac{\theta}{2} \mathbb{1} - i \sin \frac{\theta}{2} U. \quad \diamond$$

Understanding single-qubit states as vectors on the Bloch sphere, the rotation operators implement a rotation of the state about the x -, y - and z -axis, respectively. This also links to the fact that the Pauli X , Y and Z operators correspond to $-iR_X(\pi)$, $-iR_Y(\pi)$ and $-iR_Z(\pi)$, respectively.

REMARK 6.5. In the computational basis with the standard ordering, these operators are represented by the matrices

$$\begin{aligned} R_X(\theta) &= \begin{pmatrix} \cos \frac{\theta}{2} & -i \sin \frac{\theta}{2} \\ -i \sin \frac{\theta}{2} & \cos \frac{\theta}{2} \end{pmatrix}, \\ R_Y(\theta) &= \begin{pmatrix} \cos \frac{\theta}{2} & -\sin \frac{\theta}{2} \\ \sin \frac{\theta}{2} & \cos \frac{\theta}{2} \end{pmatrix}, \\ R_Z(\theta) &= \begin{pmatrix} e^{-i\frac{\theta}{2}} & 0 \\ 0 & e^{i\frac{\theta}{2}} \end{pmatrix}. \end{aligned} \quad \diamond$$

The rotation operators are important when we want to prepare an arbitrary qubit state starting from the computational basis $|0\rangle$ state:

PROPOSITION 6.6. *Every single-qubit state $|v\rangle \in \mathbb{C}^2$ can be written (up to equivalence by a global phase) as*

$$|v\rangle \sim R_Z(\phi)R_Y(\theta)|0\rangle$$

for some $\theta \in [0, \pi]$ and $\phi \in [0, 2\pi)$.

PROOF. We calculate

$$R_Z(\phi)R_Y(\theta)|0\rangle = e^{-i\frac{\phi}{2}} \left(\cos \frac{\theta}{2} |0\rangle + e^{i\phi} \sin \frac{\theta}{2} |1\rangle \right)$$

and see that this is equivalent to (3.2). \square

Furthermore, the rotation operators allow us to express an arbitrary single-qubit unitary operator:

PROPOSITION 6.7. *Every unitary operator U on \mathbb{C}^2 can be written as*

$$U = e^{i\lambda} R_Z(\phi)R_Y(\theta)R_Z(\psi)$$

for some $\lambda, \phi, \psi \in [0, 2\pi)$, $\theta \in [0, \pi]$.

PROOF. We calculate (using the matrix representation in the computational basis)

$$e^{i\lambda} R_Z(\phi) R_Y(\theta) R_Z(\psi) = \begin{pmatrix} e^{i(\lambda - \frac{\psi}{2} - \frac{\phi}{2})} \cos \frac{\theta}{2} & -e^{i(\lambda + \frac{\psi}{2} - \frac{\phi}{2})} \sin \frac{\theta}{2} \\ e^{i(\lambda - \frac{\psi}{2} + \frac{\phi}{2})} \sin \frac{\theta}{2} & e^{i(\lambda + \frac{\psi}{2} + \frac{\phi}{2})} \cos \frac{\theta}{2} \end{pmatrix}. \quad (6.1)$$

Now we show that an arbitrary qubit unitary has the above form. For an arbitrary operator

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

to be unitary, we need

$$\mathbb{1} = UU^\dagger = \begin{pmatrix} |a|^2 + |b|^2 & a\bar{c} + b\bar{d} \\ c\bar{a} + d\bar{b} & |c|^2 + |d|^2 \end{pmatrix} = U^\dagger U = \begin{pmatrix} |a|^2 + |c|^2 & \bar{a}b + \bar{c}d \\ \bar{b}a + \bar{d}c & |b|^2 + |d|^2 \end{pmatrix},$$

which implies (by comparing the diagonal entries)

$$|a|^2 + |b|^2 = |c|^2 + |d|^2 = |a|^2 + |c|^2 = |b|^2 + |d|^2 = 1.$$

Therefore, we can parametrise the matrix elements as

$$a = e^{i\alpha} \cos \xi, \quad b = -e^{i\beta} \sin \xi, \quad c = e^{i\gamma} \sin \xi, \quad d = e^{i\delta} \cos \xi. \quad (6.2)$$

for some angles $\alpha, \beta, \gamma, \delta \in [0, 2\pi)$ and $\xi \in [0, \frac{\pi}{2}]$. From the off-diagonal entries, we obtain

$$(e^{i(\alpha-\gamma)} - e^{i(\beta-\delta)}) \cos \xi \sin \xi = 0,$$

and therefore $\alpha - \gamma = \beta - \delta$. Let us introduce the angles

$$\begin{aligned} \theta &= 2\xi \in [0, \pi] \\ \lambda &= \frac{\alpha + \delta}{2} = \frac{\beta + \gamma}{2} \in [0, 2\pi) \\ \psi &= \beta - \alpha = \delta - \gamma \in [0, 2\pi) \\ \phi &= \gamma - \alpha = \delta - \beta \in [0, 2\pi) \end{aligned}$$

and therefore

$$\xi = \frac{\theta}{2}, \quad \alpha = \lambda - \frac{\psi}{2} - \frac{\phi}{2}, \quad \beta = \lambda + \frac{\psi}{2} - \frac{\phi}{2}, \quad \gamma = \lambda - \frac{\psi}{2} + \frac{\phi}{2}, \quad \delta = \lambda + \frac{\psi}{2} + \frac{\phi}{2},$$

which satisfy $\alpha, \beta, \gamma, \delta \in [0, 2\pi)$, $\xi \in [0, \frac{\pi}{2}]$ and $\alpha - \gamma = \beta - \delta$. Substituting these angles into Equation (6.2), we recover the form in Equation (6.1). \square

In order to be able to perform an arbitrary multi-qubit unitary, single-qubit rotations are clearly not sufficient. One also needs entangling operations, and the following theorem shows that the CNOT gate with arbitrary single-qubit unitaries suffices:

THEOREM 6.8. *Any unitary U on an n -qubit system $(\mathbb{C}^2)^{\otimes n}$ can be implemented by composing CNOT gates with single-qubit unitaries.*

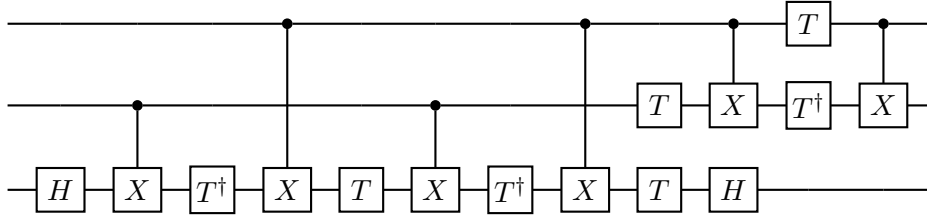


FIGURE 6.4. Decomposition of the Toffoli gate using CNOT gates, Hadamard gates and T gates.

The proof of this theorem is beyond the scope of this module, but it can be shown using the singular value decomposition and the cosine-sine decomposition. Together with Proposition 6.7, this theorem implies that every n -qubit unitary can be written as a product of CNOT gates, rotations around the Y and Z axis, and phase shifts (i.e. operators of the form $e^{i\lambda\mathbf{1}}$).

EXAMPLE 6.9. The Toffoli gate can be decomposed in terms of CNOT gates, Hadamard gates and $T = e^{i\frac{\pi}{8}}R_Z(\frac{\pi}{4})$ gates as shown in Figure 6.4. Any other decomposition using CNOT gates also requires at least 6 of them.

6.5 UNIVERSAL QUANTUM COMPUTER

A quantum computer can be thought of as a machine that implements some quantum circuits. A quantum computer is called *universal*, if it can implement arbitrary quantum circuits, and therefore can—in theory—perform arbitrary computation on n -qubit states by implementing an arbitrary unitary operation. In the section above, we saw that a relatively simple set of gates (rotations and a phase) can be used to implement arbitrary n -qubit gates. That is, by composing elements of this set, one can build a universal quantum computer. As such, universality is also a property of sets of gates:

DEFINITION 6.10. A set of quantum gates is *exactly universal* if any unitary operator can be constructed as a finite circuit of its elements. \diamond

Theorem 6.8 implies that the set consisting of the CNOT gate and all single-qubit gates is exactly universal. In fact, it can also be shown that any entangling two-qubit gate can replace the CNOT gate (recall from Proposition 4.26, that the only *non*-entangling two-qubit gates are the SWAP gate and all gates of the form $U \otimes V$).

Exact universality is quite a challenging feat. However, for practical applications, it is sufficient if we are able to *approximate* arbitrary n -qubit unitaries:

DEFINITION 6.11. A set of quantum gates S is *universal* if any unitary operator U on $(\mathbb{C}^2)^{\otimes n}$ can be approximated arbitrarily well as a finite circuit of its elements. That is, for any accuracy $\varepsilon > 0$ there exists a circuit U_m consisting of m gates from S such that

$$\sup_{|v\rangle} \|(U - U_m)|v\rangle\| \leq \varepsilon,$$

where the supremum is over all states $|v\rangle \in (\mathbb{C}^2)^{\otimes n}$. \diamond

So, what type of gates are required to get a universal set of quantum gates? Some necessary properties are the following:

1. The set of gates must have the *ability to create superpositions* (of the computational basis elements). For example, the CNOT gate cannot create superpositions, but the Hadamard gate can.
2. The set of gates must have the *ability to create entanglement*. Since no single-qubit gate can create entanglement, at least one ‘entangling’ multiple-qubit gate must be included.
3. The set of gates must have the *ability to create complex phases* (in the computational basis). For example, the Z gate cannot create a complex phase but the S gate, defined as $S := |0\rangle\langle 0| + i|1\rangle\langle 1|$, can.

These properties are clearly necessary but they are not sufficient. For example, the set $\{C_X, H, S\}$ satisfies the three properties but all possible compositions of these operators yields only a discrete subgroup of all possible unitary operators.¹³

THEOREM 6.12. *The following sets of gates are universal:*

1. $\{C_X, H, T\}$,
2. $\{C_S, H\}$,
3. $\{\text{TOFFOLI}, H, S\}$.

It is useful to realise that the sets above are not particularly special. For example, almost any two-qubit gate can be used to replace C_X in the set 1, and almost any single-qubit gate can be used to replace H in the sets 1–3.

When we have a universal set of gates, it is important to figure out if it is practical to approximate other gates using gates from that set.

THEOREM 6.13 (SOLOVAY–KITAEV). *Let \mathcal{S} be any universal set of gates that is closed under inversion.¹⁴ There exists a constant $c \lesssim 3.97$ such that any unitary can be approximated to accuracy $\varepsilon > 0$ with $O(\log^c(\frac{1}{\varepsilon}))$ gates from \mathcal{S} .¹⁵*

The growth rate in this theorem is good enough that practical implementations of quantum algorithms using only a limited set of gates are not slowed down so much that they lose their advantage over a classical algorithm. For example, a growth rate of the form $O(\frac{c}{\varepsilon})$ could make Grover’s algorithm, described in the next chapter, impractical.

¹³ Circuits of these gates can be efficiently simulated using classical computers, a result known as the Gottesman–Knill theorem.

¹⁴ That is, if $G \in \mathcal{S}$, then also its inverse $G^{-1} = G^\dagger \in \mathcal{S}$.

¹⁵ The big O notation $f(x) = O(g(x))$ means that there exists a positive number M and a real number x_0 such that $|f(x)| \leq Mg(x)$ for all $x \geq x_0$.

7 Quantum algorithms

In this section, we look at some quantum circuits that implement some computational tasks. We start off with one of the core ideas in quantum computation, called *quantum parallelism*.

7.1 QUANTUM PARALLELISM

At the core of the quantum algorithms that we will consider throughout this section is a unitary operator U_f that implements a ‘quantum version’ of a function f . Normally we consider functions from bit-strings to bits, that is, $f : \mathbb{Z}_2^n \rightarrow \mathbb{Z}_2$, where $n \in \mathbb{N}$ and $\mathbb{Z}_2 = \{0, 1\}$ with the modulo 2 addition (also called exclusive OR, or XOR) $\oplus : \mathbb{Z}_2 \times \mathbb{Z}_2 \rightarrow \mathbb{Z}_2$,

$$\begin{aligned}0 \oplus 0 &= 0 \\0 \oplus 1 &= 1 \\1 \oplus 0 &= 1 \\1 \oplus 1 &= 0,\end{aligned}$$

and the multiplication

$$\begin{aligned}0 \cdot 0 &= 0 \\0 \cdot 1 &= 0 \\1 \cdot 0 &= 0 \\1 \cdot 1 &= 1.\end{aligned}$$

For two bit strings $x, y \in \mathbb{Z}_2^n$, we also use the notation $x \oplus y$ for the bit-wise XOR, and $x \cdot y$ for the bit-wise product. Later we will use that

$$\sum_{x \in \mathbb{Z}_2^n} (-1)^{x \cdot x} = 0 \quad \text{and} \quad \sum_{x \in \mathbb{Z}_2^n} (-1)^{x \cdot y} = \begin{cases} 2^n & \text{if } y = 0, \\ 0 & \text{otherwise.} \end{cases}$$

As input to U_f , we will typically take a superposition of all possible inputs (a superposition of all bit-strings $x = (x_{n-1}, \dots, x_0) \in \mathbb{Z}_2^n$ encoded in an n -qubit state, see a couple of paragraphs later) and therefore, in some sense, we “compute the function f for all possible values simultaneously”. This is called *quantum parallelism*.

Quantum parallelism on its own brings no advantage over a classical algorithm. A measurement performed directly after applying U_f to the superposition will just randomly return one of the possible outcomes and

thus nothing is gained. However, by making smart use of the state of the system after quantum parallelism and U_f , one can sometimes, as we shall see, build quantum algorithms that are superior to classical algorithms.

To create a superposition of all possible input values, we use the *Hadamard–Walsh gate* $W := H \otimes H \otimes \cdots \otimes H$, which is just a tensor product of Hadamard gates. For an n -qubit system, we have

$$\begin{aligned} W |00 \dots 0\rangle &= (H \otimes H \otimes \cdots \otimes H) |00 \dots 0\rangle \\ &= \frac{1}{\sqrt{2^n}} (|0\rangle + |1\rangle) \otimes (|0\rangle + |1\rangle) \otimes \cdots \otimes (|0\rangle + |1\rangle) \\ &= \frac{1}{\sqrt{2^n}} (|0 \dots 00\rangle + |0 \dots 01\rangle + |0 \dots 10\rangle + \cdots + |1 \dots 11\rangle) \\ &= \frac{1}{\sqrt{2^n}} \sum_{x=0}^{2^n-1} |x\rangle, \end{aligned}$$

where, in the last step, we switched from binary to decimal notation. The state is therefore an equal superposition of all the inputs encoded into qubits, $|x\rangle = |x_{n-1}\rangle \otimes \cdots \otimes |x_0\rangle$.

EXAMPLE 7.1. We can make use of this notation, e.g. when computing the action of the Hadamard–Walsh gate W on a general n -qubit computational basis state. For $x = (x_{n-1}, \dots, x_0) \in \mathbb{Z}_2^n$, we have

$$\begin{aligned} W |x\rangle &= (H \otimes \cdots \otimes H) (|x_{n-1}\rangle \otimes \cdots \otimes |x_0\rangle) \\ &= \frac{1}{\sqrt{2^n}} (|0\rangle + (-1)^{x_{n-1}} |1\rangle) \otimes \cdots \otimes (|0\rangle + (-1)^{x_0} |1\rangle) \\ &= \frac{1}{\sqrt{2^n}} \sum_{y \in \mathbb{Z}_2^n} (-1)^{x_{n-1}y_{n-1}} |y_{n-1}\rangle \otimes \cdots \otimes (-1)^{x_0y_0} |y_0\rangle \\ &= \frac{1}{\sqrt{2^n}} \sum_{y \in \mathbb{Z}_2^n} (-1)^{x \cdot y} |y\rangle, \end{aligned}$$

since

$$(-1)^{x_j y_j} |y_j\rangle = \begin{cases} |0\rangle & \text{if } y_j = 0, \\ (-1)^{x_j} |1\rangle & \text{if } y_j = 1. \end{cases} \quad \diamond$$

As indicated above, we implement functions as certain unitary operators. If $f : \mathbb{Z}_2^n \rightarrow \mathbb{Z}_2$, it is implemented as the unitary operator U_f on $n+1$ qubits, defined in the computational basis as

$$U_f(|x\rangle \otimes |y\rangle) = |x\rangle \otimes |y \oplus f(x)\rangle \quad \longleftrightarrow \quad \begin{array}{c} |x\rangle \xrightarrow{\quad n \quad} \boxed{U_f} |x\rangle \\ |y\rangle \text{ ————— } \boxed{U_f} |y \oplus f(x)\rangle \end{array}$$

where $x \in \mathbb{Z}_2^n$ and $y \in \mathbb{Z}_2$. Together with the application of the Hadamard–Walsh gate, this gate gives

$$U_f((W |00 \dots 0\rangle) \otimes |0\rangle) = \frac{1}{\sqrt{2^n}} U_f \sum_{x \in \mathbb{Z}_2^n} |x\rangle \otimes |0\rangle = \frac{1}{\sqrt{2^n}} \sum_{x \in \mathbb{Z}_2^n} |x\rangle \otimes |f(x)\rangle.$$

That is, the superposition contains all of the 2^n function values of f entangled with their corresponding input values. As described above, this is called quantum parallelism.

EXAMPLE 7.2. The Toffoli gate from the previous chapter (denoted by CC_X , referring to the two control bits) can be interpreted as computing the AND (conjunction) of two bits (also denoted by $x \wedge y$ for two bits x and y):

$$|x_1\rangle \otimes |x_0\rangle \otimes |0\rangle \mapsto |x_1\rangle \otimes |x_0\rangle \otimes |x_1 \wedge x_0\rangle.$$

To use quantum parallelism, we combine it with the Hadamard–Walsh gate

$$(W|00\rangle) \otimes |0\rangle = \frac{1}{2}(|000\rangle + |010\rangle + |100\rangle + |110\rangle)$$

to obtain

$$CC_X((W|00\rangle) \otimes |0\rangle) = \frac{1}{2}(|000\rangle + |010\rangle + |100\rangle + |111\rangle). \quad \diamond$$

The superposition in the result can be read as the truth table for AND:

x_1	x_0	$x_1 \wedge x_0$
0	0	0
0	1	0
1	0	0
1	1	1

There is a pervasive misconception that quantum parallelism allows an exponential speedup compared to a classical algorithm, because the superposition contains the exponential number of 2^n functional values. However, any measurement in the computational basis can only extract one input-output pair, so it still takes (at least) 2^n evaluations of U_f to obtain all pairs.

The art in quantum computing sits in carefully extracting information from the superposition obtained by quantum parallelism. There are two standard strategies:

1. Find global properties of the set of function values as, e.g. in the Deutsch–Jozsa algorithm.
2. Amplify the probability for measuring values of interest as, e.g. in Grover’s algorithm.

7.2 DEUTSCH’S ALGORITHM

Suppose we are given a black-box function (also called an *oracle*) $f : \mathbb{Z}_2 \rightarrow \mathbb{Z}_2$, i.e. an unknown function that we can learn about by ‘calling’ it: giving it an input and reading its output. Classically, to decide whether f is constant ($f(0) = f(1)$) or non-constant ($f(0) \neq f(1)$) requires exactly two calls to the oracle—one needs to evaluate both $f(0)$ and $f(1)$. Deutsch’s algorithm

shows that the analogous quantum mechanical problem can be solved with only one call to the oracle.

In the quantum case, the oracle f is given by the unitary operator U_f

$$U_f(|x\rangle \otimes |y\rangle) = |x\rangle \otimes |y \oplus f(x)\rangle \quad \longleftrightarrow \quad \begin{array}{ccc} |x\rangle & \text{---} & \boxed{U_f} & \text{---} & |x\rangle \\ & & & & \\ |y\rangle & \text{---} & & & |y \oplus f(x)\rangle \end{array}$$

where $x, y \in \mathbb{Z}_2$. The action of this gate on the state $|+\rangle \otimes |-\rangle$ is

$$\begin{aligned} U_f(|+\rangle \otimes |-\rangle) &= \frac{1}{2} U_f((|0\rangle + |1\rangle) \otimes (|0\rangle - |1\rangle)) \\ &= \frac{1}{2} |0\rangle \otimes (|f(0)\rangle - |f(0) \oplus 1\rangle) + \frac{1}{2} |1\rangle \otimes (|f(1)\rangle - |f(1) \oplus 1\rangle) \\ &= \frac{1}{2} ((-1)^{f(0)} |0\rangle + (-1)^{f(1)} |1\rangle) \otimes (|0\rangle - |1\rangle) \\ &= \begin{cases} (-1)^{f(0)} |+\rangle \otimes |-\rangle, & \text{if } f(0) = f(1), \\ (-1)^{f(0)} |-\rangle \otimes |-\rangle, & \text{if } f(0) \neq f(1), \end{cases} \end{aligned}$$

where in the third equality we used that

$$|f(x)\rangle - |f(x) \oplus 1\rangle = (-1)^{f(x)}(|0\rangle - |1\rangle).$$

We therefore see that a measurement of the first qubit of $U_f(|+\rangle \otimes |-\rangle)$ in the Hadamard basis yields the outcome corresponding to the state $|+\rangle$ with probability

$$p = \begin{cases} 1 & \text{if } f(0) = f(1), \\ 0 & \text{if } f(0) \neq f(1). \end{cases}$$

That is, we learn with certainty whether f is constant or non-constant after only one call to the oracle U_f . The complete circuit for Deutsch's algorithm is a special case of that for the Deutsch–Jozsa algorithm shown in Figure 7.1 with $n = 1$.

7.3 DEUTSCH–JOZSA ALGORITHM

The Deutsch–Jozsa algorithm generalises Deutsch's algorithm to n qubits. More concretely, we have access to an oracle $f : \mathbb{Z}_2^n \rightarrow \mathbb{Z}_2$ that is either constant ($f(x) = 0$ or $f(x) = 1$ for all $x \in \mathbb{Z}_2^n$) or balanced ($f(x) = 0$ for half of $x \in \mathbb{Z}_2^n$ and $f(x) = 1$ for the other half). Our task is to determine which of the two possibilities are true. A classical algorithm needs 2 calls to the oracle in the best case and $2^{n-1} + 1$ in the worst case. The Deutsch–Jozsa algorithm shows that the analogous quantum mechanical problem can be solved with only one call to the oracle.

As for Deutsch's algorithm, the oracle f is given by the unitary operator U_f

$$U_f(|x\rangle \otimes |y\rangle) = |x\rangle \otimes |y \oplus f(x)\rangle \quad \longleftrightarrow \quad \begin{array}{ccc} |x\rangle & \text{---}^n & \boxed{U_f} & \text{---} & |x\rangle \\ & & & & \\ |y\rangle & \text{---} & & & |y \oplus f(x)\rangle \end{array}$$

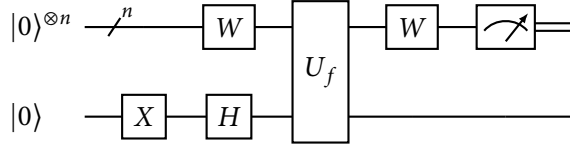


FIGURE 7.1. Circuit for the Deutsch-Jozsa algorithm.

where $x \in \mathbb{Z}_2^n$ and $y \in \mathbb{Z}_2$. The rest of the circuit is also a direct generalisation, and it is shown in Figure 7.1.

The state is prepared by applying $H^{\otimes n} \otimes (HX)$ on the state $|0\rangle^{\otimes n} \otimes |0\rangle$:

$$|0\rangle^{\otimes n} \otimes |0\rangle \mapsto \frac{1}{\sqrt{2^n}} \sum_{x \in \mathbb{Z}_2^n} |x\rangle \otimes |-\rangle.$$

The oracle U_f transforms this to

$$\frac{1}{\sqrt{2^{n+1}}} \sum_{x \in \mathbb{Z}_2^n} |x\rangle \otimes (|f(x)\rangle - |f(x) \oplus 1\rangle) = \frac{1}{\sqrt{2^n}} \sum_{x \in \mathbb{Z}_2^n} (-1)^{f(x)} |x\rangle \otimes |-\rangle.$$

Moreover, we have

$$(|0\rangle\langle 0| H)^{\otimes n} = (|0\rangle\langle +|)^{\otimes n} = \frac{1}{\sqrt{2^n}} \sum_{y \in \mathbb{Z}_2^n} |0\rangle^{\otimes n} \langle y|,$$

so that

$$\begin{aligned} (|0\rangle\langle 0| H)^{\otimes n} \frac{1}{\sqrt{2^n}} \sum_{x \in \mathbb{Z}_2^n} (-1)^{f(x)} |x\rangle &= \frac{1}{2^n} \sum_{y \in \mathbb{Z}_2^n} \sum_{x \in \mathbb{Z}_2^n} (-1)^{f(x)} \langle y|x\rangle |0\rangle^{\otimes n} \\ &= \frac{1}{2^n} \sum_{x \in \mathbb{Z}_2^n} (-1)^{f(x)} |0\rangle^{\otimes n}. \end{aligned}$$

Therefore, a measurement of the first n qubits in the Hadamard basis (or, equivalently, in the computational basis after an application of the Hadamard gate) yields the outcome corresponding to the state $|+\rangle^{\otimes n}$ (or, equivalently, the outcome corresponding to the state $|0\rangle^{\otimes n}$ if measuring in the computational basis after applying $H^{\otimes n}$) with probability

$$p = \left| \frac{1}{2^n} \sum_{x \in \mathbb{Z}_2^n} (-1)^{f(x)} \right|^2 = \begin{cases} 1 & \text{if } f \text{ is constant,} \\ 0 & \text{if } f \text{ is balanced.} \end{cases}$$

As before, we learn with certainty whether f is constant or balanced after only one call to the oracle U_f , vastly outperforming any possible classical algorithm.

7.4 SUBROUTINE: CHANGE PHASE OF SUBSET

In the following, we will look at *Grover's algorithm*, a quantum search algorithm. This algorithm uses a subroutine, changing the phase of a subset. First, we look at this subroutine.

AIM. Change the phase of terms in a superposition $|v\rangle = \sum_{x \in \mathbb{Z}_2^n} \alpha_x |x\rangle$ depending on whether x is in a set $X \subset \mathbb{Z}_2^n$ or not. That is, we wish to implement the unitary operator

$$S_X : |v\rangle = \sum_{x \in \mathbb{Z}_2^n} \alpha_x |x\rangle \mapsto - \sum_{x \in X} \alpha_x |x\rangle + \sum_{x \notin X} \alpha_x |x\rangle.$$

ASSUMPTION. We are given a(n efficiently computable) function $f : \mathbb{Z}_2^n \rightarrow \mathbb{Z}_2$, where

$$f(x) = \begin{cases} 1 & \text{if } x \in X, \\ 0 & \text{otherwise,} \end{cases}$$

which is implemented as unitary operator U_f (as in the Deutsch–Jozsa algorithm).

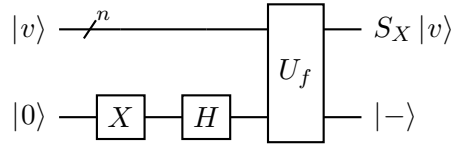


FIGURE 7.2. Circuit for S_X .

CIRCUIT. A circuit to achieve S_X is shown in Figure 7.2. That this circuit is sufficient to affect the desired phase change for the set X is seen from the calculation

$$\begin{aligned} U_f(|v\rangle \otimes |- \rangle) &= U_f \sum_{x \in X} \alpha_x |x\rangle \otimes |- \rangle + U_f \sum_{x \notin X} \alpha_x |x\rangle \otimes |- \rangle \\ &= - \sum_{x \in X} \alpha_x |x\rangle \otimes |- \rangle + \sum_{x \notin X} \alpha_x |x\rangle \otimes |- \rangle \\ &= (S_X |v\rangle) \otimes |- \rangle, \end{aligned}$$

where we used that

$$U_f(|x\rangle \otimes |- \rangle) = \begin{cases} -|x\rangle \otimes |- \rangle & \text{if } x \in X \\ |x\rangle \otimes |- \rangle & \text{if } x \notin X. \end{cases}$$

7.5 GROVER'S ALGORITHM

Consider a function $f : \mathbb{Z}_2^n \rightarrow \mathbb{Z}_2$ such that

$$f(x) = \begin{cases} 1 & \text{if } x = \omega, \\ 0 & \text{otherwise,} \end{cases}$$

for some fixed, but unknown, $\omega \in \mathbb{Z}_2^n$. The aim is to find ω . This is in general a difficult task, but once ω is found it should be simple to verify, which is what we mean by f being efficiently computable (it's easy to compute $f(x)$ for a given x). In the language of the previous section, we are given f as an oracle.

Finding ω classically takes on average 2^{n-1} calls to f . Grover's algorithm shows that the analogous quantum mechanical problem can be solved faster, but differently from the Deutsch–Jozsa algorithm the speed-up is only quadratic.

FIRST IDEA. Use quantum parallelism to consider all potential solutions ‘simultaneously’. That is, build the algorithm around the state

$$\begin{aligned} |s\rangle &:= H^{\otimes n} |0\rangle^{\otimes n} = \frac{1}{\sqrt{2^n}} \sum_{x \in \mathbb{Z}_2^n} |x\rangle = \frac{1}{\sqrt{2^n}} |\omega\rangle + \sqrt{\frac{2^n - 1}{2^n}} |s'\rangle \\ &= \sin \frac{\theta}{2} |\omega\rangle + \cos \frac{\theta}{2} |s'\rangle, \end{aligned}$$

where we used

$$|s'\rangle := \frac{1}{\sqrt{2^n - 1}} \sum_{x \neq \omega} |x\rangle, \quad \text{and} \quad \sin \frac{\theta}{2} = \frac{1}{\sqrt{2^n}}.$$

It is useful to think of $|s\rangle$ as a vector in the two-dimensional space spanned by $|\omega\rangle$ and $|s'\rangle$.

SECOND IDEA. Note that a measurement of $|s\rangle$ in the computational basis has the outcome ω with probability $\sin^2 \frac{\theta}{2} = \frac{1}{2^n}$, which is exponentially small in n . To increase the probability of finding ω , the underlying idea is to rotate $|s\rangle$ towards $|\omega\rangle$. We perform this rotation in two steps, and in order to demonstrate these steps, we assume that the system is in some generic state $|v\rangle = \cos \phi |s'\rangle + \sin \phi |\omega\rangle$ (initially $\phi = \frac{\theta}{2}$, but this angle will change throughout the algorithm, as we will see in the following). The two steps on this generic state are:

1. Change the sign of the amplitude of $|\omega\rangle$ using the subroutine $S_\omega = S_{\{\omega\}}$ developed in Section 7.4 (note that this uses the oracle U_f):

$$|v\rangle = \cos \phi |s'\rangle + \sin \phi |\omega\rangle \mapsto S_\omega |v\rangle = \cos \phi |s'\rangle - \sin \phi |\omega\rangle =: |v'\rangle.$$

Geometrically, $|v\rangle$ is reflected about the hyperplane perpendicular to $|\omega\rangle$, see Figure 7.3. Note that $S_\omega = \mathbb{1} - 2|\omega\rangle\langle\omega|$.

2. Amplify the amplitude of $|\omega\rangle$ by reflection about the hyperplane perpendicular to ‘the average’ $|s\rangle$ using the operator

$$U_s = 2|s\rangle\langle s| - \mathbb{1}.$$

Applied to $|v'\rangle = \cos \phi |s'\rangle - \sin \phi |\omega\rangle$, it yields

$$\begin{aligned} U_s |v'\rangle &= 2|s\rangle\langle s|v'\rangle - |v'\rangle = 2\langle s|(\cos \phi |s'\rangle - \sin \phi |\omega\rangle)|s\rangle - |v'\rangle \\ &= \left(2 \cos \phi \cos \frac{\theta}{2} - 2 \sin \phi \sin \frac{\theta}{2}\right) |s\rangle - \cos \phi |s'\rangle + \sin \phi |\omega\rangle \\ &= \cos(\phi + \theta) |s'\rangle + \sin(\phi + \theta) |\omega\rangle. \end{aligned}$$

The effect of this operation is illustrated in Figure 7.3.

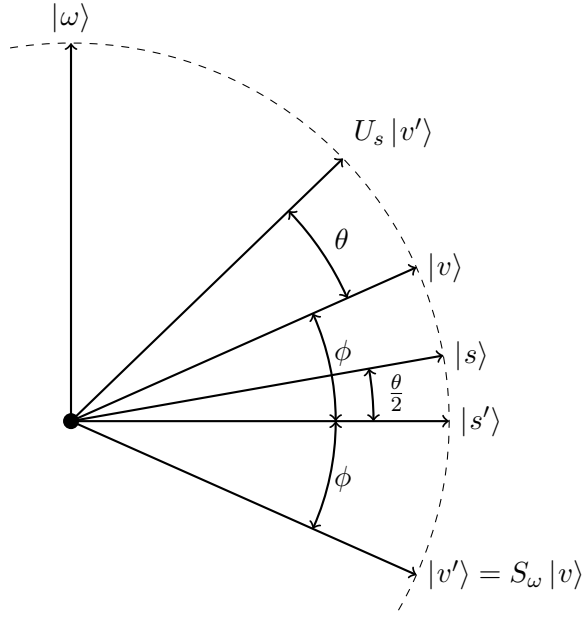


FIGURE 7.3. Illustration of the vectors and angles involved in one step of Grover's algorithm.

Note that since usually θ is relatively small, the first few steps increase the amplitude of $|\omega\rangle$ (that is, $\sin(\phi + \theta) > \sin(\phi)$ in the first few steps) and decrease the amplitude of $|s'\rangle$ (that is, $\cos(\phi + \theta) < \cos(\phi)$ in the first few steps). In particular, after the first step, $\sin(\frac{\theta}{2} + \theta) > \sin \frac{\theta}{2}$.

IMPLEMENTATION. Repeating the steps described above, one can stepwise amplify the amplitude of $|\omega\rangle$ and thereby increase the probability of the corresponding outcome of finding the correct solution. Namely, we have

$$(U_s S_\omega)^k |s\rangle = \cos\left(\frac{\theta}{2} + k\theta\right) |s'\rangle + \sin\left(\frac{\theta}{2} + k\theta\right) |\omega\rangle$$

The probability to obtain the outcome ω in a measurement of this state in the computational basis is

$$p_\omega = \sin^2\left(\frac{\theta}{2} + k\theta\right).$$

Hence the number of iterations k is optimal when

$$\sin\left(\frac{\theta}{2} + k\theta\right) \approx 1 \quad \Leftrightarrow \quad k \approx \frac{\pi}{2\theta} - \frac{1}{2}.$$

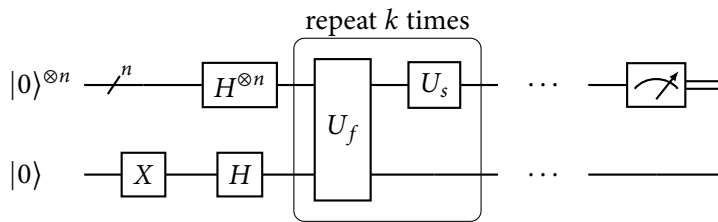


FIGURE 7.4. Circuit for Grover's algorithm.

Using the small-angle approximation,

$$\frac{1}{\sqrt{2^n}} = \sin \frac{\theta}{2} \approx \frac{\theta}{2}$$

and thus

$$k \approx \frac{\pi}{4} \sqrt{2^n} - \frac{1}{2} \approx \frac{\pi}{4} \sqrt{2^n}.$$

That is, this algorithm finds ω with high probability after $k = O(\sqrt{2^n})$ calls to the oracle U_f . This is quadratically better than any classical algorithm, but the speed-up is less than what one might expect after seeing the dramatic speed-up that the Deutsch–Jozsa algorithm can achieve. However, it can be shown that (asymptotically) there is no faster quantum algorithm for this problem than Grover’s algorithm.

In most cases there is no k such that $p_\omega = 1$ and there is no certainty that this algorithm actually finds ω . Therefore, as the final step, we need to check (classically) if the outcome is the solution ω . Otherwise we have to repeat the algorithm (but usually after repeating the algorithm a few times, ω is found, and the number of repetitions doesn’t make the algorithm less useful than the classical algorithm for large n).