Part I

Introduction to Quantum Information Science

Chapter 1

The State-Vector Formalism

The goal of quantum mechanics is to provide a mathematical framework that provides the means to rigorous describe and predict the behavior of certain physical objects, typically particles like electrons, photons etc. In contrast to a typical physics textbook on quantum mechanics, here we are not interested in analyzing *specific* physical systems, like the hydrogen atom; instead, we want to understand the *general behavior* of "quantum-mechanical objects", and how this — often strange — behavior affects the concepts of *computation* and *information* as we know them, and which arose from abstracting the behavior of typical (and thus non-quantum) information-processing devices. This angle of quantum mechanics is called **quantum computing** or **quantum information-processing** if the focus lies on aspects related more to computing, or **quantum information theory** if the focus lies on aspects related to the behavior of information. In its union, it is referred to as **quantum information science**.

In this section, we introduce the so-called state vector formalism of quantum mechanics, which is one particular framework for describing "quantum-mechanical objects" and their behavior. The state-vector formalism is somewhat limited in that there are certain aspects that can not be (well) captured; however, those are not very relevant in the context of quantum information-processing, which is what we focus on.

1.1 State Spaces and State Vectors

Let $\mathcal{H}$ be an arbitrary Hilbert space.

Definition 1.1. *$\mathcal{S}(\mathcal{H})$ denotes the set of all norm-1 vectors in $\mathcal{H}$, i.e.,*

$$\mathcal{S}(\mathcal{H}) := \{|\varphi\rangle \in \mathcal{H} \,|\, \||\varphi\rangle\| = 1\}.$$

A vector $|\varphi\rangle \in \mathcal{S}(\mathcal{H})$ is called a **state vector**.

The connection to physics is as follows. Any "quantum-mechanical object", which we abstract for the time being by the pictogram ⚛, is associated to a Hilbert space $\mathcal{H}$, called the **state space** of ⚛. The **(quantum) state** of ⚛, which is meant to determine the future behavior of ⚛, can then be mathematically described by a state vector $|\varphi\rangle \in \mathcal{S}(\mathcal{H})$. In the context of quantum information science, such an abstract "quantum-mechanical object" ⚛ is also referred to as a **(quantum) system** or, in the context of quantum computing, as a **register**. However, we tend to be a bit sloppy with the terminology and do not always distinguish well between the quantum system, the state of the system, and the description of the state by means of a state vector. Later on, when we consider multiple quantum systems, we will refer to them by A, B

etc. rather than ⚛, and their respective state spaces are then by default denoted by $\mathcal{H}_A, \mathcal{H}_B$ etc.

In case of a 2-dimensional state space $\mathcal{H}$, which may then be assumed to be $\mathcal{H} = \mathbb{C}^2$, the quantum system ⚛, respectively the state vector $|\varphi\rangle \in \mathcal{S}(\mathcal{H})$ describing its state, is typically called a **qubit**, and in case of dimension $d > 2$ it is sometimes referred to as a **qu*d*it**.

Let $\{|i\rangle\}_{i\in I}$ be some fixed orthonormal basis of $\mathcal{H}$. Then, any state vector $|\varphi\rangle \in \mathcal{S}(\mathcal{H})$ can be written as a linear combination, we also say: as a **superposition**,

$$|\varphi\rangle = \sum_i \alpha_i |i\rangle$$

of the $|i\rangle$'s, where the α_i's, called **amplitudes**, satisfy

$$\sum_i |\alpha_i|^2 = \sum_{ij} \overline{\alpha}_i \alpha_j \langle i|j\rangle = \langle\varphi|\varphi\rangle = 1\,.$$

In case of a 2-dimensional state space $\mathcal{H}$, we consider a fixed orthonormal basis $\{|0\rangle, |1\rangle\}$ of $\mathcal{H}$, called the **computational basis** (or Z**-basis** or **rectilinear basis**). In case $\mathcal{H} = \mathbb{C}^2$, which we may well assume without loss of generality, the computional basis is given by the canonical basis

$$|0\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \qquad \text{and} \qquad |1\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

A state vector $|\varphi\rangle \in \mathcal{S}(\mathbb{C}^2)$ can then be written as a superposition $|\varphi\rangle = \alpha_0|0\rangle + \alpha_1|1\rangle$ of $|0\rangle$ and $|1\rangle$, with $\alpha_0, \alpha_1 \in \mathbb{C}$ such that $|\alpha_0|^2 + |\alpha_1|^2 = 1$ (see Figure 1.1).

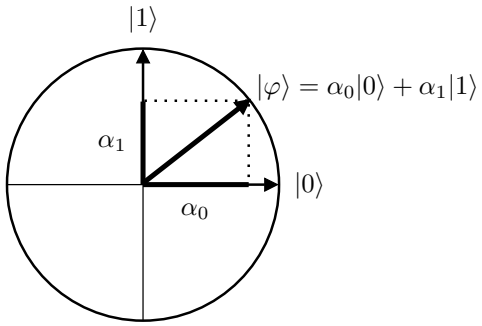


Figure 1.1: A qubit with (real) amplitudes α_0 and α_1.

Another orthonormal basis of the qubit state space $\mathcal{H} = \mathbb{C}^2$ that is important to us is the so-called **Hadamard basis** (or X**-basis** or **diagonal basis**), given by the two basis vectors

$$|+\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \end{pmatrix} = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$$

and

$$|-\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ -1 \end{pmatrix} = \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle)\,.$$

For a state space $\mathcal{H}$ with arbitrary dimension d, the **computational** basis is denoted by $|0\rangle, |1\rangle, \ldots, |d-1\rangle$.[1] We will later see possible generalizations of the Hadamard basis to higher dimensions.

[1] Sometimes, it is more convenient to write the computational basis as $|1\rangle, |2\rangle, \ldots, |d\rangle$ instead.

Strictly speaking, we can — and will — also consider $\mathcal{S}(\mathbb{C}) = \{\omega \in \mathbb{C} \,|\, |\omega| = 1\}$, but then we do not think of or refer to the elements as (state) vectors; instead, an element of $\mathcal{S}(\mathbb{C})$ is called a **phase**. We point out that, together with the multiplication in $\mathbb{C}$, $\mathcal{S}(\mathbb{C})$ forms a group.

Definition 1.2. *Two state vectors $|\varphi\rangle, |\varphi'\rangle \in \mathcal{S}(\mathcal{H})$ are* **equivalent**, *denoted as $|\varphi\rangle \equiv |\varphi'\rangle$, if $|\varphi\rangle = \omega|\varphi'\rangle$ for some $\omega \in \mathcal{S}(\mathbb{C})$.*

We will see that two state vectors that are equivalent, or, as we also say, *equal up to a global phase*, behave identically under the physically-relevant operations that are of interest to us, and thus they describe the same state.

1.2 Unitary Evolution

The natural operation U to apply to a state vector $|\varphi\rangle \in \mathcal{S}(\mathcal{H})$, so as to transform it into another state vector $|\varphi'\rangle = U|\varphi\rangle \in \mathcal{S}(\mathcal{H})$, is a *unitary* $U \in \mathcal{U}(\mathcal{H})$; indeed, unitarity ensures that the norm is preserverd. From the physics perspective, for any unitary $U \in \mathcal{U}(\mathcal{H})$ there exists a way to manipulate a quantum system ⚛ with state space $\mathcal{H}$ so that the (possibly unknown) state $|\varphi\rangle$ of ⚛ evolves from $|\varphi\rangle$ to $|\varphi'\rangle = U|\varphi\rangle$. Vice versa, any physical manipulation of a given system ⚛ *without causing it to interact with the environment* corresponds to a unitary $U \in \mathcal{U}(\mathcal{H})$.

A particular unitary that we have already (implicitly) encountered is the **Hadamard operator** $H \in \mathcal{U}(\mathbb{C}^2)$, which maps the computational basis into the Hadamard basis, i.e.,

$$H : |0\rangle \mapsto |+\rangle, |1\rangle \mapsto |-\rangle$$

and thus maps any $|\varphi\rangle = \alpha_0|0\rangle + \alpha_1|1\rangle \in \mathcal{S}(\mathbb{C}^2)$ to

$$H|\varphi\rangle = \alpha_0 H|0\rangle + \alpha_1 H|1\rangle = \frac{\alpha_0}{\sqrt{2}}\big(|0\rangle + |1\rangle\big) + \frac{\alpha_1}{\sqrt{2}}\big(|0\rangle - |1\rangle\big) = \frac{\alpha_0 + \alpha_1}{\sqrt{2}}|0\rangle + \frac{\alpha_0 - \alpha_1}{\sqrt{2}}|1\rangle\,.$$

H is Hermitian, i.e. $H^\dagger = H$, and self-inverse, i.e., $H^2 = \mathbb{I}$, and so it also maps the Hadamard basis back into the computational basis.

Other important examples are the **Pauli operators** (or "gates") $X, Y, Z \in \mathcal{U}(\mathbb{C}^2)$, which act as

$$X : |0\rangle \mapsto |1\rangle, |1\rangle \mapsto |0\rangle\,, \quad Y : |0\rangle \mapsto i|1\rangle, |1\rangle \mapsto -i|0\rangle \quad \text{and} \quad Z : |0\rangle \mapsto |0\rangle, |1\rangle \mapsto -|1\rangle\,.$$

As matrices (with respect to the computational basis) they are

$$X = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \quad , \qquad Y = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix} \qquad \text{and} \qquad Z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}\,.$$

Like H, they are Hermitian and self-inverse; as a matter of fact, they satisfy

$$X^2 = Y^2 = Z^2 = -iXYZ = iZYX = \mathbb{I}\,.$$

In particular, $\{\mathbb{I}, \pm X, \pm iX, \pm Y, \pm iY, \pm Z, \pm iZ\}$ forms a group, the (1-qubit) **Pauli group**.

Another noteworthy example of a (parameterized) 1-qubit unitary is the **phase shift gate**

$$S_\theta : |0\rangle \mapsto |0\rangle, |1\rangle \mapsto e^{i\theta}|1\rangle$$

for an arbitrary $\theta \in \mathbb{R}$, which changes the state by means of a **local** phase. Important special cases of the phase shift gate are $S_\pi = Z$, $S_{\pi/2}$, which is called **phase gate** and denoted by S, and $S_{\pi/4}$, which is sometimes denoted by T and then referred to as ***T*-gate**.[2]

More generally, one may consider *isometries* V, which have the defining property that $V^\dagger V = \mathbb{I}$ (but not necessarily $VV^\dagger = \mathbb{I}$). An isometry $V \in \mathcal{L}(\mathcal{H}, \mathcal{H}')$ maps a state vector $|\varphi\rangle \in \mathcal{S}(\mathcal{H})$ into a state vector $|\varphi'\rangle = V|\varphi\rangle \in \mathcal{S}(\mathcal{H}')$ in a "bigger" Hilbert space, i.e., where $\dim(\mathcal{H}') \geq \dim(\mathcal{H})$. Physically, an isometry is an embedding of a quantum system into a bigger one.

1.3 Measurements

Another (physically-relevant) way to act on a state is by means of a *measurement*, formally captured by means of the following definitions.

Definition 1.3. *Let I be an arbitrary non-empty finite set. A family $\mathbf{M} = \{M_i\}_{i\in I}$ of operators $M_i \in \mathcal{L}(\mathcal{H})$ with*

$$\sum_{i\in I} M_i^\dagger M_i = \mathbb{I}$$

is called a **family of measurement operators***, or simply a* **measurement***. The set of all such families for a given index set I is denoted $\mathcal{M}eas_I(\mathcal{H})$.*

Definition 1.4. *For any $\mathbf{M} \in \mathcal{M}eas_I(\mathcal{H})$ and any state vector $|\varphi\rangle \in \mathcal{S}(\mathcal{H})$, we define for every $i \in I$:*

$$p_i := \|M_i|\varphi\rangle\|^2 = \langle\varphi|M_i^\dagger M_i|\varphi\rangle \qquad \text{and} \qquad |\varphi^i\rangle := \frac{1}{\sqrt{p_i}} M_i|\varphi\rangle \in \mathcal{S}(\mathcal{H})\,,$$

with $|\varphi^i\rangle$ undefined in case $p_i = 0$.

Note that $p_i \geq 0$ by definition, and $\sum_i p_i = 1$ by the defining property of a measurement and the fact that $|\varphi\rangle$ has norm 1. So, the p_i's form a *probability distribution*. Also, we see here that changing the phase of the state vector $|\varphi\rangle$ has no effect on the p_i's.

Here, the physical relevance is as follows, referred to as **Born's rule**. Any measurement device, which interacts with a quantum system ⚛ and produces a measurement outcome (like a number on a monitor), is described by a measurement $\mathbf{M} \in \mathcal{M}eas_I(\mathcal{H})$, where $\mathcal{H}$ is the state space of ⚛. Using this device to measure a particular ⚛ with (possibly unknown) state $|\varphi\rangle$ then has the effect that outcome $i \in I$ is *observed* (i.e. displayed on the monitor) with probability p_i, and the state of ⚛ *collapses* to the **post-measurement state** $|\varphi^i\rangle$.

Motivated by the above, for given state vector $|\varphi\rangle \in \mathcal{S}(\mathcal{H})$ and measurement $\mathbf{M} = \{M_i\}_{i\in I}$, we can — and will — speak of "the probability to observe (a particular outcome) i", which is then well defined to be $p_i = \langle\varphi|M_i^\dagger M_i|\varphi\rangle$, or "the probability that the measurement outcome lies in the set T", which is defined as $\sum_{i\in T} p_i$, etc.

We can push this further and for instance considering yet another measurement $\mathbf{N} = \{N_j\}_{j\in J}$ that is applied to the post-measurement state that results from the first measurement; we may then speak of "the probability to observe i in the first measurement and j in the second", which is well defined to be

$$p_{ij} := p_i\, \langle\varphi^i|N_j^\dagger N_j|\varphi^i\rangle = \langle\varphi|M_i^\dagger N_j^\dagger N_j M_i|\varphi\rangle\,.$$

Here, $p_{j|i} := \langle\varphi^i|N_j^\dagger N_j|\varphi^i\rangle$ can be naturally understood as "the probability of the second measurement producing observation j *conditioned* on the first producing i".

[2] Confusingly, $S_{\pi/4}$ is sometimes also referred to as $\pi/8$ gate; the reason for this is that, up to an unimportant global phase $e^{i\pi/8}$, it is equal to the diagonal matrix with $e^{\pm i\pi/8}$ on its diagonal.

The above also shows that the considered sequential application of the measurements $\mathbf{M}$ and $\mathbf{N}$ has "the same effect" as the one measurement $\{N_j M_i\}_{(i,j)\in I\times J}$, meaning that both induce the same (join) probabilities p_{ij} (and the same post-measurement states $|\varphi^{ij}\rangle$). We leave it as a simple exercise to verify that $\{N_j M_i\}_{(i,j)\in I\times J}$ satisfies Definition 1.3.

Similar results hold when composing a measurement with a unitary operator: we leave it as an exercise to show that a unitary followed by a measurement has the same effect as one suitably chosen measurement, and the same for a measurement followed by a unitary (that may then depend on the measurement outcome).

1.4 Projective Measurements

In this section, we introduce a special yet important class of measurements.

Definition 1.5. $\mathbf{M} = \{M_i\}_{i\in I} \in \mathcal{M}eas_I(\mathcal{H})$ *is called a* **projective** *(or* **Von Neumann***) measurement if M_i is a projection for every $i \in I$. Furthermore,* $\mathbf{M}$ *is called a* **rank-1 projective** *measurement if every M_i is of the form $M_i = |e_i\rangle\langle e_i|$ with $|e_i\rangle \in \mathcal{S}(\mathcal{H})$.*

The following in particular implies that a rank-1 projective measurement may be described by an orthonormal basis of $\mathcal{H}$.

Lemma 1.1. *If $\{P_i\}_{i\in I}$ is a projective measurement, then the projections P_i are pairwise mutually orthogonal: $P_i P_j = 0$ for $i \neq j$. In particular, if $\{P_i\}_{i\in I}$ is a rank-1 projective measurement, and thus $P_i = |e_i\rangle\langle e_i|$ for all $i \in I$, then $\{|e_i\rangle\}_{i\in I}$ is an orthonormal basis of $\mathcal{H}$.*

Proof. Given that $\sum_i P_i^\dagger P_i = \mathbb{I}$ and using the defining properties of projections, we see that for any $j \in I$ and $|\varphi\rangle \in \mathcal{H}$,

$$\langle\varphi|P_j|\varphi\rangle = \langle\varphi|P_j^\dagger P_j|\varphi\rangle = \sum_i \langle\varphi|P_j^\dagger P_i^\dagger P_i P_j|\varphi\rangle = \langle\varphi|P_j|\varphi\rangle + \sum_{i\neq j} \langle\varphi|P_j^\dagger P_i^\dagger P_i P_j|\varphi\rangle$$

and the claim follows from the observation that $\langle\varphi|P_j^\dagger P_i^\dagger P_i P_j|\varphi\rangle = \|P_i P_j|\varphi\rangle\|^2 \geq 0$. □

In the case of such a projective measurement $\mathbf{M} = \{P_i\}_{i\in I}$, Born's rule obviously simplifies to $p_i = \langle\varphi|P_i^\dagger P_i|\varphi\rangle = \langle\varphi|P_i|\varphi\rangle$. In case of a rank-1 projective measurement $\mathbf{M} = \{|i\rangle\langle i|\}_{i\in I}$, we say that "we measure the quantum system *in the basis* $\{|i\rangle\}_{i\in I}$"; here, Born's rule simplifies to

$$p_i = \langle\varphi||i\rangle\langle i||i\rangle\langle i||\varphi\rangle = \langle\varphi|i\rangle\langle i|\varphi\rangle = |\langle i|\varphi\rangle|^2$$

and

$$|\varphi^i\rangle = \frac{1}{\sqrt{p_i}}|i\rangle\langle i||\varphi\rangle = \frac{1}{\sqrt{p_i}}|i\rangle\langle i|\varphi\rangle \equiv |i\rangle$$

Therefore, when writing the state vector $|\varphi\rangle$ as a superposition

$$|\varphi\rangle = \sum_i \alpha_i|i\rangle$$

then the p_i's can easily be obtained from the amplitues as

$$p_i = |\alpha_i|^2 .$$

For example, let us consider the rank-1 projective measurement $\mathbf{M} = \{|0\rangle\langle 0|, |1\rangle\langle 1|\}$ given by the computational basis $\{|0\rangle, |1\rangle\}$ of $\mathcal{H} = \mathbb{C}^2$. Then, for $|\varphi\rangle = |0\rangle$ we see immediately that

$p_0 = 1$ and $p_1 = 0$, i.e., the outcome 0 is observed with certainty. Similarly, for $|\varphi\rangle = |1\rangle$ we have $p_0 = 0$ and $p_1 = 1$. Whereas for $|\varphi\rangle = |+\rangle$ and $|\varphi\rangle = |-\rangle$, we get $p_0 = p_1 = \frac{1}{2}$. Correspondingly, when we consider the rank-1 projective measurement given by the Hadamard basis $\{|+\rangle, |-\rangle\}$ and the states $|+\rangle, |-\rangle, |0\rangle, |1\rangle$.

We conclude by remarking that in the literature, one also finds the terminology that an **observable** M is measured of the system A, where M is a Hermitian matrix in $\mathcal{L}(\mathcal{H})$. In our terminology, this corresponds to the projective measurement $\mathbf{M} = \{M_i\}_{i\in I}$ where the M_i's are the orthogonal projections into the eigenspaces of M, and the i's are the corresponding (real) eigenvalues. Vice versa, every projective measurement $\mathbf{M} = \{M_i\}_{i\in I}$ (with $I \subset \mathbb{R}$) may be phrased in terms of an observable M. We do not make use of this terminology/formalism.

1.5 POVMs

In cases where one is only interested in the measurement outcome (and its distribution) but not in the post-measurement state, the general measurement formalism of a family $\mathbf{M} = \{M_i\}_{i\in I}$ of measurement matrices can be simplified.

Definition 1.6. *Let $\mathbf{E} = \{E_i\}_{i\in I}$ be a non-empty finite family of matrices $E_i \in \mathcal{L}(\mathcal{H})$. $\mathbf{E}$ is called a* **POVM** *(which stands for a "Positive-Operator Valued Measure") if*

$$E_i \geq 0 \;\; \forall\, i \in I \quad \textit{and} \quad \sum_{i\in I} E_i = \mathbb{I}\,.$$

For a finite index set I, we let $\mathcal{POVM}_I(\mathcal{H})$ denote the set of POVM's $\mathbf{E} = \{E_i\}_{i\in I}$.

The following is a trivial observation. For every measurement $\mathbf{M} = \{M_i\}_{i\in I}$ in $\mathcal{M}eas_I(\mathcal{H})$, the family $\mathbf{E} = \{E_i\}_{i\in I}$ with $E_i = M_i^\dagger M_i$ is in $\mathcal{POVM}_I(\mathcal{H})$, and

$$p_i = \langle\varphi|M_i^\dagger M_i|\varphi\rangle = \langle\varphi|E_i|\varphi\rangle$$

for all $|\varphi\rangle \in \mathcal{S}(\mathcal{H})$ and $i \in I$. Hence, every measurement gives rise to a POVM, and the POVM is sufficient to compute p_i. Vice versa, every POVM arises from some measurement:

Lemma 1.2. *For every $\mathbf{E} = \{E_i\}_{i\in I} \in \mathcal{POVM}_I(\mathcal{H})$ there exists a measurement $\mathbf{M} = \{M_i\}_{i\in I}$ such that $E_i = M_i^\dagger M_i$ for every $i \in I$.*

The existence of a decomposition $E_i = M_i^\dagger M_i$ follows immediately from the spectral decomposition (Theorem 0.3) of E_i and the positivity of E_i. For instance, $M_i := \sqrt{E_i}$, defined according to Definition 0.1, does the job. However, we stress that the decomposition $E_i = M_i^\dagger M_i$ is *not* unique in general, and therefore the post-measurement state $|\varphi^i\rangle$ is not uniquely determined by $\mathbf{E}$ (when given $|\varphi\rangle$), but p_i is. As such, the POVM formalism is applicable if we are merely interested in the measurement statistics but not in the post-measurement state.

1.6 Perfect Distinguishability

We consider the following question. If an "experimenter" is given one or another state, how easy or hard is it for him to find out in which state of the two it is by means of performing an arbitrary measurement. This motivates the following definition.

Definition 1.7. *Two state vectors $|\varphi\rangle, |\psi\rangle \in \mathcal{S}(\mathcal{H})$ are called* **perfectly distinguishable** *if there exists a POVM $\mathbf{E} = \{E_0, E_1\} \in \mathcal{POVM}_{\{0,1\}}(\mathcal{H})$ such that $\langle\varphi|E_0|\varphi\rangle = 1 = \langle\psi|E_1|\psi\rangle$.*

Theorem 1.3. *Two state vectors $|\varphi\rangle \in \mathcal{S}(\mathcal{H})$ and $|\psi\rangle \in \mathcal{S}(\mathcal{H})$ are perfectly distinguishable if and only if they are orthogonal, i.e., $\langle\varphi|\psi\rangle = 0$.*

Proof. That orthogonality is a sufficient condition for perfectly distinguishability is obvious: we simply take a measurement that is described by an orthonormal basis that contains $|\varphi\rangle$ and $|\psi\rangle$ as basis vectors, or, more formally, we set

$$E_0 = |\varphi\rangle\langle\varphi| \qquad \text{and} \qquad E_1 = \mathbb{I} - |\varphi\rangle\langle\varphi| \,.$$

Also, it should be clear that it is a necessary condition when restricting to *projective* measurements. For general measurements, we can argue as follows. Using that $E_1 \geq 0$ can be written as $E_1 = M_1^\dagger M_1$, we see that

$$0 = \langle\varphi|E_1|\varphi\rangle = \langle\varphi|M_1^\dagger M_1|\varphi\rangle = \|M_1|\varphi\rangle\|^2$$

and thus $M_1|\varphi\rangle = 0$, and hence also $E_1|\varphi\rangle = 0$. Similarly, $\langle\psi|E_0|\psi\rangle = 0$ implies $E_0|\psi\rangle = 0$. It follows that

$$\langle\varphi|\psi\rangle = \langle\varphi|(E_0 + E_1)|\psi\rangle = \langle\varphi|(E_0 + E_1^\dagger)|\psi\rangle = \langle\varphi|E_0|\psi\rangle + \langle\varphi|E_1^\dagger|\psi\rangle = 0 \,.$$

Hence, $|\varphi\rangle$ and $|\psi\rangle$ must be orthogonal. □

We now see that two states, given by state vectors $|\varphi\rangle, |\psi\rangle \in \mathcal{S}(\mathcal{H})$, are perfectly distinguishable if and only if $\langle\varphi|\psi\rangle = 0$, and they are perfectly *in*distinguishable (in the obvious sense) if and only if $|\varphi\rangle$ and $|\psi\rangle$ are identical up to the phase, i.e., $|\varphi\rangle \equiv |\psi\rangle$, or, equivalenty, $|\langle\varphi|\psi\rangle| = 1$. Thus, we understand the extreme cases. For the cases in-between, the following seems to be a suitable measure for capturing how far away we are from one or the other extreme case.

Definition 1.8. *The* **fidelity** *of two state vectors $|\varphi\rangle, |\psi\rangle \in \mathcal{S}(\mathcal{H})$ is defined as*

$$F(|\varphi\rangle, |\psi\rangle) := |\langle\varphi|\psi\rangle| \,.$$

Indeed, it turns out that the **distingushing advantage** of two states $|\varphi\rangle, |\psi\rangle \in \mathcal{S}(\mathcal{H})$, defined as

$$adv(|\varphi\rangle, |\psi\rangle) := \max_{0 \leq E_0 \leq \mathbb{I}} \big(\langle\varphi|E_0|\varphi\rangle - \langle\psi|E_0|\psi\rangle\big) = \max_{0 \leq E_0 \leq \mathbb{I}} \langle\varphi|E_0|\varphi\rangle + \langle\psi|(\mathbb{I} - E_0)|\psi\rangle - 1 \,,$$

is determined by the fidelity:

$$adv(|\varphi\rangle, |\psi\rangle) = \sqrt{1 - F(|\varphi\rangle, |\psi\rangle)^2} \,.$$

The fidelity should be thought of as a measure of distance, but obviously it is not a metric in the mathematical sense; in particular, small fidelity means that the states are far away, and large fidelity (i.e., a fidelity close to 1) means that the states are close to each other. The distingushing advantage, however, turns out to be a metric, the so called **trace distance**.

1.7 The Bloch Sphere

The Bloch sphere, which we introduce here, offers a nice geometrical description of the space of qubits *modulo the (irrelevant) phase*. First, we observe that, for any Hilbert space $\mathcal{H}$, the mapping

$$\mathcal{S}(\mathcal{H}) \to \mathcal{L}(\mathcal{H}) \,, |\varphi\rangle \mapsto |\varphi\rangle\langle\varphi|$$

induces an *injection* on $\mathcal{S}(\mathcal{H})/\!\equiv$, i.e., the set of equivalence classes $\{\omega|\varphi\rangle \,|\, \omega \in \mathcal{S}(\mathbb{C})\}$.

Given that the global phase of a state vector is irrelevant, $\rho := |\varphi\rangle\langle\varphi|$ can thus be understood as a description of the state in terms of an operator, which, in contrast to the state-vector description, is *unique*; indeed, this is how states are described in the so-called **density operator** formalism. Note that by construction, ρ is positive-semidefinite (and thus Hermitian) and has trace $\mathrm{tr}(\rho) = \mathrm{tr}(|\varphi\rangle\langle\varphi|) = \langle\varphi|\varphi\rangle = 1$.

We now focus on $\mathcal{H} = \mathbb{C}^2$. It is not too hard to see that $\{\mathbb{I}, X, Y, Z\}$ forms an $\mathbb{R}$-basis of the Hermitian operators in $\mathcal{L}(\mathbb{C}^2)$, and therefore $\rho = |\varphi\rangle\langle\varphi|$ must be of the form

$$\rho = \frac{1}{2}\big(\mathbb{I} + xX + yY + zZ\big)\,,$$

for real-valued $x, y, z \in \mathbb{R}$. Additionally, the fact that $\rho^2 = |\varphi\rangle\langle\varphi|\varphi\rangle\langle\varphi| = \rho$ implies that

$$\begin{aligned}\frac{1}{2}\big(\mathbb{I} + xX + yY + zZ\big) &= \frac{1}{4}\big(\mathbb{I} + xX + yY + zZ\big)^2 \\ &= \frac{1}{4}\big(\mathbb{I} + x^2\mathbb{I} + y^2\mathbb{I} + z^2\mathbb{I} + 2xX + 2yY + 2zZ + xy\{X,Y\} + xz\{X,Z\} + yz\{Y,Z\}\big) \\ &= \frac{1}{4}\big(1 + x^2 + y^2 + z^2\big)\mathbb{I} + \frac{1}{2}\big(xX + yY + zZ\big)\,,\end{aligned}$$

and therefore $x^2 + y^2 + z^2 = 1$. Thus, we obtain an injective mapping

$$\mathcal{S}(\mathcal{H})/\!\equiv \;\to \{(x,y,z) \in \mathbb{R}^3 \,|\, x^2 + y^2 + z^2 = 1\}$$

into the real 2-sphere, which is then referred to as **Bloch sphere**. We will also see that this mapping is also surjective. Thus, we can identify qubit states with points on the Bloch sphere.

From the observation that

$$|0\rangle\langle 0| = \begin{bmatrix}1 & 0\\ 0 & 0\end{bmatrix} = \frac{1}{2}(\mathbb{I} + Z) \qquad \text{and} \qquad |1\rangle\langle 1| = \begin{bmatrix}0 & 0\\ 0 & 1\end{bmatrix} = \frac{1}{2}(\mathbb{I} - Z)$$

we see that the computational basis vectors $|0\rangle$ and $|1\rangle$ correspond to the points $z = \pm 1$ on the Block sphere (see Figure 1.2). Similarly, the Hadamard basis vectors $|+\rangle$ and $|-\rangle$ correspond to $x = \pm 1$, while the points $y = \pm 1$ represent the vectors from the **circular basis**, given by

$$|\circlearrowleft\rangle = \frac{1}{\sqrt{2}}(|0\rangle + i|1\rangle) \qquad \text{and} \qquad |\circlearrowright\rangle = \frac{1}{\sqrt{2}}(|0\rangle - i|1\rangle)\,.$$

We can now study the action of single-qubit unitaries on the Block sphere. For instance, from the basic properties of the Pauli operators (like $XZX = -Z$ etc.), we immediately see that

$$X|\varphi\rangle\langle\varphi|X = \frac{1}{2}X\big(\mathbb{I} + xX + yY + zZ\big)X = \frac{1}{2}\big(\mathbb{I} + xX - yY - zZ\big)\,,$$

which, e.g., shows that X maps $|0\rangle$ to $|1\rangle$ and vice versa (which of course we already knew), maps $|\circlearrowleft\rangle$ to $|\circlearrowright\rangle$ and vice versa, and leaves $|+\rangle$ and $|-\rangle$ untouched (modulo the phase!). Thus, as an action on the Block sphere, X is a rotation by 180° around the axis given by $|+\rangle$ and $|-\rangle$. Correspondingly for Y and Z.

More generally, we consider the following unitaries, one for each Pauli operator and parameterized by $\theta \in \mathbb{R}$.

$$R_X(\theta) := \cos\big(\tfrac{\theta}{2}\big)\mathbb{I} - i\sin\big(\tfrac{\theta}{2}\big)X = \begin{bmatrix}\cos(\frac{\theta}{2}) & -i\sin(\frac{\theta}{2})\\ -i\sin(\frac{\theta}{2}) & \cos(\frac{\theta}{2})\end{bmatrix}$$

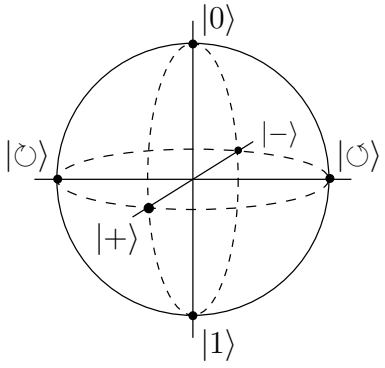


Figure 1.2: The Bloch sphere.

$$R_Y(\theta) := \cos\tfrac{\theta}{2}\mathbb{I} - i\sin\tfrac{\theta}{2}Y = \begin{bmatrix} \cos(\frac{\theta}{2}) & -\sin(\frac{\theta}{2}) \\ \sin(\frac{\theta}{2}) & \cos(\frac{\theta}{2}) \end{bmatrix}$$
$$R_Z(\theta) := \cos\tfrac{\theta}{2}\mathbb{I} - i\sin\tfrac{\theta}{2}Z = \begin{bmatrix} e^{-i\theta/2} & 0 \\ 0 & e^{i\theta/2} \end{bmatrix}.$$

It is easy to see that these are indeed unitaries with $R_{X/Y/Z}(\theta)^\dagger = R_{X/Y/Z}(-\theta)$ and

$$\begin{aligned} R_Z(\theta) Z R_Z(\theta)^\dagger &= Z \quad , \\ R_Z(\theta) X R_Z(\theta)^\dagger &= \cos(\theta)X + \sin(\theta)Y \quad \text{and} \\ R_Z(\theta) Y R_Z(\theta)^\dagger &= -\sin(\theta)X + \cos(\theta)Y\,. \end{aligned}$$

Therefore, for any $|\varphi\rangle \in \mathcal{S}(\mathbb{C}^2)$ with Bloch-sphere coordinates (x, y, z), we have

$$\begin{aligned} R_Z(\theta)|\varphi\rangle\langle\varphi|R_Z(\theta)^\dagger &= \frac{1}{2}\big(\mathbb{I} + xR_Z(\theta)XR_Z(\theta)^\dagger + yR_Z(\theta)YR_Z(\theta)^\dagger + zR_Z(\theta)ZR_Z(\theta)^\dagger\big) \\ &= \frac{1}{2}\big(\mathbb{I} + (x\cos(\theta) - y\sin(\theta))X + (x\sin(\theta) + y\cos(\theta))Y + zZ\big)\,, \end{aligned}$$

and thus $R_Z(\theta)|\varphi\rangle$ has Bloch-sphere coordinates $(x\cos(\theta) - y\sin(\theta), x\sin(\theta) + y\cos(\theta), z)$. In other words, $R_Z(\theta)$ acts as a *rotation* around the z axis, and similarly for $R_X(\theta)$ and $R_Y(\theta)$. Because of this, these are called **rotation operators**. Also, it is easy to see that

$$ZR_Z(\theta)Z = R_Z(\theta) \quad , \qquad XR_Z(\theta)X = R_Z(-\theta) \quad \text{and} \qquad YR_Z(\theta)Y = R_Z(-\theta)$$

and similarly for $R_X(\theta)$ and $R_Y(\theta)$.

We conclude with the following characterization theorem for single-qubit unitaries.

Theorem 1.4 (Z-Y decomposition)**.** *For any $U \in \mathcal{U}(\mathbb{C}^2)$ there exist $\alpha, \beta, \gamma, \delta$ such that*

$$U = e^{i\alpha}R_Z(\beta)R_Y(\gamma)R_Z(\delta)\,.$$

Proof. It is easy to see that *up to the phases of the entries*, U must be of the form

$$\begin{bmatrix} \cos(\gamma/2) & -\sin(\gamma/2) \\ \sin(\gamma/2) & \cos(\gamma/2) \end{bmatrix} = R_Y(\gamma)$$

for some $\gamma \in \mathbb{R}$. Furthermore, with the goal to remove phases, we can find $\alpha, \beta, \delta \in \mathbb{R}$ so that

$$e^{-i\alpha}R_Z(-\delta)\,U R_Z(-\beta) = \begin{bmatrix} \cos(\gamma/2) & -\sin(\gamma/2) \\ \sin(\gamma/2) & \omega\cos(\gamma/2) \end{bmatrix}$$

for some $\omega \in \mathcal{S}(\mathbb{C})$. However, for the right hand side to be unitary, which it must be if U is, ω must be 1, and thus the right hand side is $R_Y(\gamma)$. This proves the claim. □

In purely mathematical terms, identifying U with its action on the Bloch sphere gives rise to a surjective homomorphism $SU(2) \to SO(3)$ from the special unitary group of degree 2 to the special orthogonal group of degree 3. The kernel of this homomorphism is $\{\pm\mathbb{I}\}$. This in turn gives rise to an isomorphism $SU(2)/\{\pm\mathbb{I}\} \leftrightarrow SO(3)$.

Chapter 2

Multipartite Quantum Systems

The formalism introduced in the previous chapter allows us to describe individual "quantum-mechanical objects" and predict their individual behavior. In this chapter, we extend the formalism so as to be able to capture *multiple* "quantum-mechanical objects" and predict their *joint* behavior. For instance, we may want to study how the respective polarizations of two photons behave in a certain experiment.

2.1 Multipartite Quantum Systems

Consider two labeled Hilbert spaces $\mathcal{H}_A$ and $\mathcal{H}_B$ (see Section 0.6) for distinct labels A and B. By default, we then understand A and B to refer to two quantum systems, i.e., to two distinct "quantum mechanical objects" ⚛ and ✯, and $\mathcal{H}_A$ and $\mathcal{H}_B$ as their respective state spaces. Following Section 0.6, the Hilbert space $\mathcal{H}_{AB}$ with label AB is then given by the tensor product

$$\mathcal{H}_{AB} = \mathcal{H}_A \otimes \mathcal{H}_B \,.$$

$\mathcal{H}_{AB}$ is then understood to be the state space of the **bipartite** quantum system that consists of the subsystems A and B. The corresponding holds for general **multipartite** systems, consisting of an arbitrary (finite) number of subsystems. The physical relevance should be clear: the state of two (or more) "quantum mechanical objects" is described by a state vector in the tensor product of the individual state spaces.

If the state of A is given by state vector $|\varphi\rangle \in \mathcal{S}(\mathcal{H}_A)$ and the state of B by $|\psi\rangle \in \mathcal{S}(\mathcal{H}_B)$ then the state of the bipartite system AB — sometimes also referred to as the **joint** state — is given by $|\Omega\rangle = |\varphi\rangle \otimes |\psi\rangle \in \mathcal{S}(\mathcal{H}_{AB})$. We refer to such a state (vector) $|\Omega\rangle$ as a **product state**.[1] We emphasize though that in general, if A and B (i.e., the two "quantum mechanical objects" of concern) were not kept in isolation but may have interacted with each other, their joint state is described by an arbitrary state vector $|\Omega\rangle$ in $\mathcal{S}(\mathcal{H}_{AB})$. In this case, i.e., if $|\Omega\rangle$ is not a product state, we say that A and B are **entangled**; entanglement is another strange phenomenon of quantum physics.

By identifying any operator $R \in \mathcal{L}(\mathcal{H}_A)$ with $R \otimes \mathbb{I} \in \mathcal{L}(\mathcal{H}_{AB})$, we naturally recover the evolution of bipartite (or multipartite) quantum systems *when acting on a subsystem*. For instance, applying a unitary $U \in \mathcal{U}(\mathcal{H}_A)$ to subsystem A of a bipartite system AB has the effect that their joint state $|\Omega\rangle \in \mathcal{S}(\mathcal{H}_{AB})$ evolves to

$$U_A\,|\Omega\rangle_{AB} = (U \otimes \mathbb{I})|\Omega\rangle \in \mathcal{S}(\mathcal{H}_{AB})\,.$$

[1] We refrain from the common terminology of calling such a vector a "*pure* tensor".

Similarly, Definition 1.4 extends to measurements $\mathbf{M} \in \mathcal{M}eas_I(\mathcal{H}_A)$ and states $|\Omega\rangle \in \mathcal{S}(\mathcal{H}_{AB})$.

In the special case of rank-1 projective measurements, we then get the following. First, we observe that by elementary properties we have that for any $|\Omega\rangle \in \mathcal{S}(\mathcal{H}_{AB})$ and orthonormal basis $\{|i\rangle\}_{i\in I}$ of $\mathcal{H}_A$, we can write

$$|\Omega\rangle = \sum_{i\in I} \alpha_i |i\rangle|\psi_i\rangle$$

with $\alpha_i \in \mathbb{C}$ and $|\psi_i\rangle \in \mathcal{S}(\mathcal{H}_B)$ for all $i \in I$, and where $\sum_i |\alpha_i|^2 = 1$. For a rank-1 projective measurement $\{|i\rangle\langle i|\}_{i\in I} \in \mathcal{M}eas_I(\mathcal{H}_A)$ given by $\{|i\rangle\}_{i\in I}$, we then see that

$$p_i = \langle\Omega|(|i\rangle\langle i| \otimes \mathbb{I}_B)|\Omega\rangle = |\alpha_i|^2$$

and

$$|\Omega^i\rangle = \frac{1}{\sqrt{p_i}}(|i\rangle\langle i| \otimes \mathbb{I}_B)|\Omega\rangle = \frac{\alpha_i}{|\alpha_i|}|i\rangle|\psi_i\rangle \equiv |i\rangle|\psi_i\rangle\,.$$

Thus, also here, as in Section 1.4, we can easily "read out" the statistics and the corresponding post-measurement states when the original state is expressed in the basis that determines the (rank-1 projective) measurement.

We point out that actions on different subsystems *commute*. For instance, for $U \in \mathcal{U}(\mathcal{H}_A)$ and $M_i \in \mathbf{M} \in \mathcal{M}eas_I(\mathcal{H}_B)$, it holds that

$$(\mathbb{I}_A \otimes M_i)(U \otimes \mathbb{I}_B)|\Omega\rangle = (U \otimes M_i)|\Omega\rangle = (U \otimes \mathbb{I}_B)(\mathbb{I}_A \otimes M_i)|\Omega\rangle\,.$$

This reads as follows. Whether we first apply U to A and then measure B, or we first measure B and then apply U to A, we get the same probability p_i to observe outcome i:

$$p_i = \langle\Omega|(U^\dagger \otimes M_i^\dagger)(U \otimes M_i)|\Omega\rangle = \langle\Omega|(\mathbb{I}_A \otimes M_i^\dagger M_i)|\Omega\rangle$$

and the same post-measurement state:

$$|\Omega^i\rangle = (U \otimes M_i)|\Omega\rangle\,.$$

2.2 Isometries and No-Cloning

The class of operators that map states of a quantum system into states of a "bigger" system are the *isometries*, as defined in Section 0.2. An important example is the isometry

$$\mathbb{I}_A \otimes |0\rangle_B : \mathcal{H}_A \to \mathcal{H}_{AB},\ |\varphi\rangle_A \mapsto |\varphi\rangle_A \otimes |0\rangle_B\,.$$

Such a $|0\rangle_B$-state that is "appended" to a given state in this manner is called an **ancilla**. Obviously, composing this (or any) isometry with a unitary results again in an isometry. Vice versa, by basic properties of isometries as discussed in Section 0.2, the following holds.

Lemma 2.1. *Any isometry $V \in \mathcal{L}(\mathcal{H}_A, \mathcal{H}_{AB})$ equals $V = U_{AB}\,(\mathbb{I}_A \otimes |0\rangle_B)$ with $U \in \mathcal{U}(\mathcal{H}_{AB})$.*

The goal of *cloning* is to turn an unknown quantum state into two copies of the original state. Clearly, if the state to be cloned is promised to be one out of two given states that are *perfectly distinguishable*, then the state *can* be (perfectly) cloned: simply perform a measurement that tells which state it is, and then prepare this state twice "from scratch". The no-cloning theorem tells us that this is *the only* case where cloning is possible.

Theorem 2.2 (No-cloning theorem)**.** *Let $V \in \mathcal{L}(\mathcal{H}_A, \mathcal{H}_A \otimes \mathcal{H}_{A'})$ be an isometry with $\mathcal{H}_A = \mathcal{H}_{A'}$, and let $|\varphi\rangle, |\psi\rangle \in \mathcal{S}(\mathcal{H}_A)$. Then, unless $\langle\varphi|\psi\rangle = 0$ or $|\varphi\rangle \equiv |\psi\rangle$, it is not possible that both*

$$V|\varphi\rangle \equiv |\varphi\rangle|\varphi\rangle \qquad \text{and} \qquad V|\psi\rangle \equiv |\psi\rangle|\psi\rangle\,.$$

Note that Theorem 2.2 considers cloning by means of an isometry, whereas above we speak of cloning in terms of measuring and preparing new states. We will later see that there is no loss of generality here.

Proof. We show the contraposition and thus assume that both equalities do hold. By taking the inner product of these two equalities, we obtain

$$\langle\varphi|\psi\rangle = \langle\varphi|V^\dagger V|\psi\rangle \equiv \langle\varphi,\varphi|\psi,\psi\rangle = \langle\varphi|\psi\rangle^2$$

from which it follows that either $\langle\varphi|\psi\rangle = 0$, and we are done, or $|\langle\varphi|\psi\rangle| = 1$. In case of the latter, by the tightness condition for Cauchy-Schwarz, $|\varphi\rangle$ and $|\psi\rangle$ must then be equal up to a scalar ω, which must then be in $\mathcal{S}(\mathbb{C})$. □

2.3 Naimark's Dilation Theorem

We show here that projective measurements are equally powerful as general measurements when allowing "pre-processing", in the sense that any general measurement $\mathbf{M}$ on a system A can be "simulated" by means of appending an ancilla system B to A, applying a unitary transformation to the joint system AB, and then doing a projective measurement, actually doing a rank-1 projective measurement on B.

Theorem 2.3 (Naimark's dilation theorem)**.** *Let $\mathbf{M} = \{M_i\}_{i\in I} \in \mathcal{M}eas_I(\mathcal{H}_A)$, and let $\{|i\rangle\}_{i\in I}$ be an orthonormal basis of $\mathcal{H}_B = \mathbb{C}^{|I|}$. Then, there exists a unitary $U \in \mathcal{U}(\mathcal{H}_{AB})$ such that for every $|\varphi\rangle \in \mathcal{S}(\mathcal{H}_A)$ and $i \in I$*

$$M_i|\varphi\rangle \otimes |i\rangle = (\mathbb{I}_A \otimes |i\rangle\langle i|)\, U|\varphi\rangle|0\rangle\,.$$

This means, every general measurement is equivalent to appending an ancilla, applying a unitary (to the joint system), and performing a rank-1 projective measurement on the ancilla (and ignoring the collapsed state of the ancilla).

Proof. Consider $V \in \mathcal{L}(\mathcal{H}_A, \mathcal{H}_{AB})$ defined by

$$V|\varphi\rangle = \sum_{i\in I} M_i|\varphi\rangle \otimes |i\rangle$$

for any $|\varphi\rangle \in \mathcal{H}_A$. It is then clear that $(\mathbb{I}_A \otimes |i\rangle\langle i|)\, V|\varphi\rangle = M_i|\varphi\rangle \otimes |i\rangle$. Furthermore, V is an isometry since for any $|\varphi\rangle, |\psi\rangle \in \mathcal{H}_A$ it holds that

$$\begin{aligned}\langle\psi|V^\dagger V|\varphi\rangle &= \sum_{i,j\in I} (\langle\psi|M_j^\dagger \otimes \langle j|)(M_i|\varphi\rangle \otimes |i\rangle) \\ &= \sum_{i,j\in I} \langle\psi|M_j^\dagger M_i|\varphi\rangle\langle j|i\rangle = \sum_{i\in I} \langle\psi|M_i^\dagger M_i|\varphi\rangle = \langle\psi|\varphi\rangle\,.\end{aligned}$$

The existence of $U \in \mathcal{U}(\mathcal{H}_{AB})$ as required then follows from Lemma 2.1. □

If one is merely interested in the measurement outcome (and its distribution), but not in the post-measurement state, then also the converse holds: the action of appending an ancilla, possibly applying a unitary transformation (to the joint system), and then performing a measurement, can be captured by means of a general measurement. Since we are not interested in the post-measurement state here, we use the POVM formalism as introduced in Section 1.5.

Proposition 2.4. *For any* $\{E_i\}_{i\in I} \in \mathcal{POVM}_I(\mathcal{H}_A \otimes \mathcal{H}_B)$ *there exists* $\{F_i\}_{i\in I} \in \mathcal{POVM}_I(\mathcal{H}_A)$ *so that for all* $|\varphi\rangle \in \mathcal{S}(\mathcal{H}_A)$ *and* $j \in I$

$$\langle\varphi|\langle 0| E_j |\varphi\rangle|0\rangle = \langle\varphi|F_j|\varphi\rangle\,.$$

Proof. Set

$$F_i := (\mathbb{I}_A \otimes \langle 0|)\, E_i\, (\mathbb{I}_A \otimes |0\rangle)\,,$$

where $\mathbb{I}_A \otimes |0\rangle : \mathcal{H}_A \to \mathcal{H}_A \otimes \mathcal{H}_B$, $|\psi\rangle \mapsto |\psi\rangle \otimes |0\rangle$ and $\mathbb{I}_A \otimes \langle 0|$ is its adjoint. Then, the claimed equality holds by construction. It remains to argue that the F_i's form a POVM. Positivity is obvious, and for completeness, we see that

$$\sum_i F_i = (\mathbb{I}_A \otimes \langle 0|)\, \mathbb{I}_{AB}\, (\mathbb{I}_A \otimes |0\rangle) = (\mathbb{I}_A \otimes \langle 0|)\, (\mathbb{I}_A \otimes \mathbb{I}_B)\, (\mathbb{I}_A \otimes |0\rangle) = \mathbb{I}_A\,,$$

which completes the proof. □

We may understand Proposition 2.4 above as another composition result, which ensures that appending an ancilla and then performing a measurement, composes to a single measurement. Together with the composition observations in Section 1.3, it now follows that any sequence of actions — composed of unitaries, appending ancillas, and measurements, possibly adaptively chosen depending on previous measurement outcomes — can be replaced by a single measurement that leads to the same distribution on the observed information.

2.4 "Quantum" versus "Classical" Information

So far, and we will to a large extent continue doing so, we have carefully distinguished between *classical* and *quantum* information. Formally, "classical information", like the outcome of a measurement, is captured by an element x of some given non-empty finite set $\mathcal{X}$. On the other hand, "quantum information" is captured by a state vector $|\varphi\rangle$ in some given Hilbert space. We want to argue here that we may also use the quantum formalism to capture classical information, i.e., in other words, we may understand quantum information as a strict generalization of classical information.

For this purpose, for any given (non-emtpy finite) set $\mathcal{X}$, we consider a fixed orthonormal basis $\{|x\rangle\}_{x\in\mathcal{X}}$ of the state space $\mathcal{H} = \mathbb{C}^{|\mathcal{X}|}$, and we identify $x \in \mathcal{X}$ with $|x\rangle \in \mathcal{S}(\mathcal{H})$. It is in this sense that the qubit states $|0\rangle$ and $|1\rangle$ represent the respective classical bits 0 and 1. We note that such an "encoding of classical information into a quantum state" can be "decoded" simply by measuring the "encoding" $|x\rangle$ in the considered basis $\{|x\rangle\}_{x\in\mathcal{X}}$: the classical measurement outcome x is observed with probability 1.

We can also use the quantum formalism to capture classical information *processing*, by identifying classical functions by unitary operators. For example, we see that the logical `not` function $\neg : \{0,1\} \to \{0,1\}$, $x \mapsto x \oplus 1$, where $\oplus$ is the addition modulo 2, is captured by the the Pauli-X unitary, as introduced in Section 1.2:

$$|\neg x\rangle = X|x\rangle$$

for any $x \in \{0,1\}$. Another example is the 2-qubit **SWAP** operator $SWAP \in \mathcal{U}(\mathbb{C}^2 \otimes \mathbb{C}^2)$, defined by

$$SWAP|x\rangle|y\rangle = |y\rangle|x\rangle$$

for all $x, y \in \{0,1\}$, which captures the 2-bit function that swaps the input bits by means of a 2-qubit unitary. Another simple yet important example is the 2-qubit **control-NOT** operator $CNOT \in \mathcal{U}(\mathbb{C}^2 \otimes \mathbb{C}^2)$, defined by

$$CNOT|x\rangle|y\rangle = |x\rangle|x \oplus y\rangle$$

for all $x, y \in \{0,1\}$, which captures the 2-bit function that applies the logic `not` to the second bit if and only if the first is 1. We emphasize that this representation of a function by means of a unitary crucially depends on the choice of basis used for representing classical information by means of quantum states. Indeed, it is interesting to see how *CNOT* acts on an input that is classical with respect to the Hadamard basis, i.e., what $CNOT(H|x\rangle \otimes H|y\rangle)$ evaluates to (when expressed in the Hadamard basis $\{H|0\rangle, H|1\rangle\}$ again). We leave this as an exercise.

We also emphasize that the above approach for representing a classical function f by means of the operator $|x\rangle \mapsto |f(x)\rangle$ only works if the function is *injective*, as otherwise the operator is not unitary. In order to deal with an arbitrary function, one uses the following approach.

Definition 2.1. *For any function $f : \mathcal{X} \to \mathcal{Y}$, we define $U_f \in \mathcal{U}(\mathcal{H}_X \otimes \mathcal{H}_Y)$ given by*

$$U_f : |x\rangle|y\rangle \mapsto |x\rangle|y \oplus f(x)\rangle$$

so that

$$U_f|x\rangle|0\rangle = |x\rangle|f(x)\rangle\,.$$

Here, it is understood that fixed bases $\{|x\rangle\}_{x\in\mathcal{X}}$ and $\{|y\rangle\}_{y\in\mathcal{Y}}$ of $\mathcal{H}_X = \mathbb{C}^{|\mathcal{X}|}$ and $\mathcal{H}_Y = \mathbb{C}^{|\mathcal{Y}|}$ have been respectively chosen, and $\oplus$ is an operation that turns $\mathcal{Y}$ into an Abelian group with neutral element 0. Often, this group structure is naturally given.

In this light, we actually have that $CNOT = U_{id}$ for the identity function $id : \{0,1\} \to \{0,1\}$, $x \mapsto x$, but this is not how we think of *CNOT*.

2.5 Control Unitaries

The control-NOT operator $CNOT \in \mathcal{U}(\mathbb{C}^2 \otimes \mathbb{C}^2)$ can also be understood in that it applies the Pauli X operator to the target qubit if (and only if) the control (qu)bit is set. This naturally generalizes.

Definition 2.2. *For any $U \in \mathcal{U}(\mathcal{H})$, the corresponding* **control unitary** *$C(U) \in \mathcal{U}(\mathbb{C}^2 \otimes \mathcal{H})$ (w.r.t. $\{|0\rangle, |1\rangle\}$) is defined as*

$$C(U) := |0\rangle\langle 0| \otimes \mathbb{I} + |1\rangle\langle 1| \otimes U\,,$$

so that $C(U)|x\rangle|\varphi\rangle = |x\rangle \otimes U^x|\varphi\rangle$ for arbitrary $x \in \{0,1\}$ and $|\varphi\rangle \in \mathcal{H}$. More generally, for arbitrary $n \in \mathbb{N}$ the **multi-control unitary** *$C^n(U) \in \mathcal{U}(\mathbb{C}^2 \otimes \cdots \otimes \mathbb{C}^2 \otimes \mathcal{H})$ is defined to map*

$$C^n(U) : |x_1\rangle \cdots |x_n\rangle|\varphi\rangle \mapsto |x_1\rangle \cdots |x_n\rangle \otimes U^{x_1 \cdots x_n}|\varphi\rangle$$

for arbitrary $x_1, \ldots, x_n \in \{0,1\}$ and $|\varphi\rangle \in \mathcal{H}$.

We note that the above definition of $C(U)$ is so that the first qubit is the **control qubit** and the second qubit is the **target state**, but we also speak of a control unitary and write $C(U)$ if it is the other way round, or, in case of a multi-control unitary, if the target state is at an arbitrary position. For instance, one could consider labeled Hilbert spaces, say using "1" and "2" as labels, and $C_1(U_2)$ would then be controlled by (the system labeled with) "1" and have (the system labeled with) "2" as target, and vice versa for $C_2(U_1)$ then. However, we typically clarify this matter in an ad-hoc manner, or by means of picturing them appropriately as **gates** in a **quantum circuit** (see the upcoming figures)

We also emphasize that even though the definition of a (multi-)control unitary is in terms of how $C^n(U)$ acts when the control qubits are classical, i.e. $|0\rangle$ or $|1\rangle$, the action of the control unitary is well defined on the entire space, and thus may be applied to an *arbitrary* state. Furthermore, maybe somewhat counterintuitive, $C^n(U)$ may in such a case then actually modify the control qubits.

Obviously, we have $CNOT = C(X)$. Furthermore, $C^2(X)$ is referred to as **Toffoli gate**. Of course, we can also consider variations of (multi-)control unitaries, where the unitary U is applied conditioned on another setting of the control qubit(s) than being (all) one. Formally, for any $c = (c_1, \ldots, c_n) \in \{0,1\}^n$ we can consider

$$C^n[c](U) := (X^{c_1 \oplus 1} \otimes \cdots \otimes X^{c_n \oplus 1} \otimes \mathbb{I})\, C^n(U)\, (X^{c_1 \oplus 1} \otimes \cdots \otimes X^{c_n \oplus 1} \otimes \mathbb{I})\,,$$

which is such that

$$C^n[c](U)|x\rangle|\varphi\rangle = \begin{cases} |x\rangle \otimes U|\varphi\rangle & \text{if } x = c \\ |x\rangle|\varphi\rangle & \text{else} \end{cases} .$$

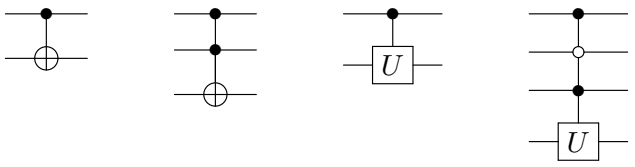


Figure 2.1: Pictorial gate representations of *CNOT*, the Toffoli gate, $C(U)$ and $C^3[1,0,1](U)$.

2.6 Decomposing Control Unitaries

Our goal here will be to show that for any *single-qubit* unitary $U \in \mathcal{U}(\mathbb{C}^2)$, the corresponding (multi-)control unitary $C^n(U)$ can be decomposed into *CNOT* and single-qubit unitaries.

First, we need the following technical result.

Lemma 2.5. *For any $U \in \mathcal{U}(\mathbb{C}^2)$ there exist $A, B, C \in \mathcal{U}(\mathbb{C}^2)$ and $\alpha \in \mathbb{R}$ so that*

$$ABC = \mathbb{I} \qquad \textit{and} \qquad e^{i\alpha} AXBXC = U\,.$$

Proof. By means of the *Z-Y* decomposition (Theorem 1.4) and introducing a factor 2 for convenience, we can write U as $U = e^{i\alpha} R_Z(2\beta) R_Y(2\gamma) R_Z(2\delta)$. Setting

$$A := R_Z(2\beta) R_Y(\gamma) \quad , \qquad B := R_Y(-\gamma) R_Z(-\beta - \delta) \qquad \text{and} \qquad C := R_Z(-\beta + \delta)\,,$$

we then immediately see that $ABC = \mathbb{I}$, but also, by basic properties of X, R_Y and R_Z,

$$AXBXC = R_Z(2\beta) R_Y(\gamma) X R_Y(-\gamma) X X R_Z(-\beta-\delta) X R_Z(-\beta+\delta) = R_Z(2\beta) R_Y(2\gamma) R_Z(2\delta)\,,$$

which proves the claim. □

Theorem 2.6. *For any $U \in \mathcal{U}(\mathbb{C}^2)$ there exist $A, B, C \in \mathcal{U}(\mathbb{C}^2)$ and $\alpha \in \mathbb{R}$ so that*

$$C(U) = (S_\alpha \otimes A)\,CNOT\,(\mathbb{I} \otimes B)\,CNOT\,(\mathbb{I} \otimes C)$$

See Figure 2.2 below for the representation of the claimed decomposition of $C(U)$ in the form of a (pictorially represented) **quantum circuit**.

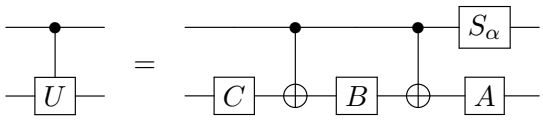


Figure 2.2: Computing $C(U)$ with *CNOT* and single qubit gates.

Remark 2.1. Note that, by convention, such a circuit acts on any input state vector by sequentially applying the **gates** starting from the *left*, while e.g. the corresponding term in Theorem 2.6 acts by sequentially applying the unitaries starting from the *right*.

Proof. Choosing the unitaries A, B, C and $\alpha \in \mathbb{C}$ as promised by Lemma 2.5, it is clear that $(\mathbb{I} \otimes A)\,CNOT\,(\mathbb{I} \otimes B)\,CNOT\,(\mathbb{I} \otimes C)$, where the phase shift gate S_α is omitted, maps

$$|0\rangle|\varphi\rangle \mapsto |0\rangle \otimes ABC|\varphi\rangle = |0\rangle|\varphi\rangle \qquad \text{and} \qquad |1\rangle|\varphi\rangle \mapsto |1\rangle \otimes AXBXC|\varphi\rangle\ .$$

It remains to show that S_α, acting on the first qubit, leaves $|0\rangle|\varphi\rangle$ untouched and maps the other into $|1\rangle \otimes e^{i\alpha} AXBXC|\varphi\rangle$, but this holds by definition of S_α. □

Arbitrary double-control unitaries can be computed from single-control unitaries as follows; the proof is left as an exercise.

Proposition 2.7. *Let $U \in \mathcal{U}(\mathcal{H})$, and let $V \in \mathcal{U}(\mathcal{H})$ be so that $V^2 = U$. Then $C^2(U)$ decomposes into CNOT, $C(V)$ and $C(V^\dagger)$ operations, as given in Figure 2.3.*

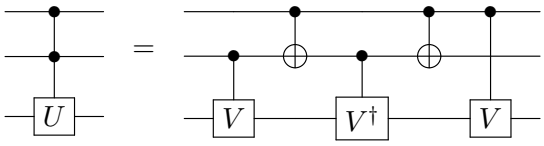


Figure 2.3: Computing $C^2(U)$ with *CNOT* and single-control unitaries.

Remark 2.2. We could circumvent the representation of the claimed decomposition by means of a quantum circuit by labelling the three qubits as $1, 2$ and 3, say, and then express the claimed operator equality as

$$C^2_{12}(U_3) = C_1(V_3)\ CNOT_{12}\ C_2(V_3^\dagger)\ CNOT_{12}\ C_2(V_3)\,.$$

However, such an expression seems harder to parse than a quantum circuit.

By replacing U with $C^{n-2}(U)$ and V with $C^{n-2}(V)$, and applying induction to compute $C\big(C^{n-2}(V)\big) = C^{n-1}(V)$, we obtain the following.

Corollary 2.8. *For any unitary $U \in \mathcal{U}(\mathcal{H})$, the multi-control unitary $C^n(U)$ decomposes into a sequence of CNOT's and control unitaries $C(V)$ with $V \in \mathcal{U}(\mathcal{H})$, all acting on $(\mathbb{C}^2)^{\otimes n} \otimes \mathcal{H}$.*

Using Theorem 2.6 to further decompose the control unitaries (in case $\mathcal{H} = \mathbb{C}^2$), we obtain the following.

Corollary 2.9. *For any single-qubit unitary $U \in \mathcal{U}(\mathbb{C}^2)$, the multi-control unitary $C^n(U)$ decomposes into CNOT's and single-qubit unitaries, all acting on $(\mathbb{C}^2)^{\otimes(n+1)}$.*

We point out that the number of gates to be computed in the above recursive construction for $C^n(U)$ is exponential in n. The following gives a more efficient way. First, note that applying Proposition 2.7 to $V \in \mathcal{U}(\mathbb{C}^2)$ with $V^2 = X$ gives us the means to compute the Toffoli gate with a single-qubit unitary and *CNOT*. Then, Figure 2.4 illustrates how a multi-control unitary $C^n(U)$ can be computed by means of the single-control unitary $C(U)$ and Toffolis, using $n-1$ "work qubits" that start off and end up again in state $|0\rangle$.

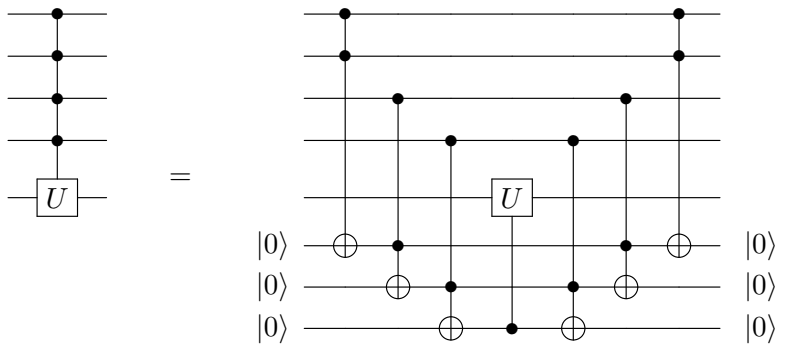


Figure 2.4: Computing $C^n(U)$ with Toffolis and $C(U)$, for the case $n = 4$.

Formally, we have the following.

Proposition 2.10. *For any $U \in \mathcal{U}(\mathcal{H})$ and for $V \in \mathcal{U}\big((\mathbb{C}^2)^{\otimes n} \otimes \mathcal{H} \otimes (\mathbb{C}^2)^{\otimes(n-1)}\big)$ defined by (the obvious generalization to an arbitrary n of) the right hand side in Figure 2.4, we have*

$$V|x\rangle|\varphi\rangle|\mathbf{0}\rangle = C^n(U)|x\rangle|\varphi\rangle \otimes |\mathbf{0}\rangle$$

for any $x \in \{0,1\}^n$ and $|\varphi\rangle \in \mathcal{H}$, and where $|\mathbf{0}\rangle = |0\rangle^{\otimes(n-1)}$.

Remark 2.3. In order for the above equality to extend to *any* $|\Omega\rangle \in (\mathbb{C}^2)^{\otimes n} \otimes \mathcal{H}$, not necessarily $|\Omega\rangle = |x\rangle|\varphi\rangle$, it is crucial that the "work qubits" end up again in state $|\mathbf{0}\rangle$, and not in something that, say, depends on x.

Chapter 3

Entanglement

Entanglement gives rise to paradoxical effects. Recall that a measurement affects the state of the system, and in the case of an entangled state measuring one subsystem will affect the state *as a whole*. It thus appears that a "quantum-mechanical object", when entangled with another one, *knows* that a measurement has been performed on the other, and what was the measurement and the outcome, even though there is no known means for such information to be communicated between the objects, which at the time of measurement may be separated by arbitrarily large distances.

Such phenomena were the subject of a 1935 paper by Albert Einstein, Boris Podolsky, and Nathan Rosen, and several papers by Erwin Schrödinger shortly thereafter, describing what came to be known as the *EPR paradox*. Einstein considered such behavior to be impossible, referring to it as "spukhafte Fernwirkung", i.e., "spooky action at a distance".

Einstein's goal with his 1935 paper was to demonstrate that the inherent non-determinism of the proposed theory of quantum mechanics leads to absurd effects. In this way, he wanted to advocate the need for a "hidden variable" explanation, which would assume measurement outcomes to be *pre-determined* yet unknown. Interestingly, in 1964 by John Bell, it was then shown how to experimentally rule out such a hidden variable explanation by means of a small twist to Einstein's EPR paradox (see Section 3.3).

Entanglement is also the crucial "resource" for **quantum telepartation**, which we briefly discuss in Section 3.2.

3.1 (Maximally) Entangled States

As already introduced, two systems are called **entangled** if the joint state is not a product state. An important entangled 2-qubit state is the so-called **EPR pair**, named after Einstein, Podolsky and Rosen, and also known as the first of the four **Bell states** (see Section 3.2):

$$|\Phi\rangle_{AB} = |\Phi^+\rangle_{AB} = \frac{1}{\sqrt{2}}\big(|0\rangle_A|0\rangle_B + |1\rangle_A|1\rangle_B\big) \in \mathcal{H}_A \otimes \mathcal{H}_B\,,$$

with $\mathcal{H}_A = \mathbb{C}^2 = \mathcal{H}_B$. It has the following peculiar property. If subsystem A (i.e., the first of the two qubits) is measured in the computational basis $\{|0\rangle, |1\rangle\}$, then $i \in \{0,1\}$ is observed with probability $p_i = \frac{1}{2}$ and the state of system B collapses to $|i\rangle$.[1] Thus, if subsequently B is measured in the computational basis, then *the same* $i \in \{0,1\}$ is observed (with probability 1). The corresponding holds when first B and then A is measured. Even more interesting, due to

[1] Indeed, $|\Phi\rangle$ is of the form $|\Phi\rangle = \sum_j \alpha_j |j\rangle \otimes |\psi_j\rangle$ with $\alpha_0 = \alpha_1 = \frac{1}{\sqrt{2}}$ and $|\psi_0\rangle = |0\rangle$ and $|\psi_1\rangle = |1\rangle$.

the fact that

$$|\Phi\rangle = \frac{1}{\sqrt{2}}\big(|+\rangle|+\rangle + |-\rangle|-\rangle\big) = \frac{1}{\sqrt{2}}\big(H|0\rangle \otimes H|0\rangle + H|1\rangle \otimes H|1\rangle\big) \in \mathcal{H}_A \otimes \mathcal{H}_B\,,$$

which is verified by a straightforward calculation, the corresponding also holds when measuring A and B in the Hadamard basis $\{|+\rangle, |-\rangle\}$. In summary, if A and B are measured in *the same* basis (computational or Hadamard) then two *fully correlated* random bits will be observed; if they are measured in "opposite" bases (one computational, and the other Hadamard) then two *completely uncorrelated* random bits will be observed.

By our terminology, a bipartite state is either entangled or not. However, we want to give a bit of an intuition what it should mean for a state to be *strongly* or *weakely* entangled; later, this will be made (more) rigorous. A state $|\Omega\rangle \in \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_B)$ is "strongly entangled" if for a suitable choice of an orthonormal basis $\{|i\rangle\}_{i\in I}$ of $\mathcal{H}_A$, writing the state under consideration as $|\Omega\rangle = \sum_{i\in I} \alpha_i |i\rangle|\psi_i\rangle$ has the property that (1) the $|\psi_i\rangle$'s are close to be *mutually orthogonal*, and (2) the distribution given by the $|\alpha_i|^2$'s is close to the *uniform distribution* over I. Obviously, for a product state, i.e., a state that is not entangled at all, for any choice of basis $\{|i\rangle\}_{i\in I}$, the $|\psi_i\rangle$'s are all identical and thus highly non-orthogonal, unless for those where $\alpha_i = 0$, which then contradicts (2). On the other hand, an EPR pair satisfies both (1) and (2) perfectly, and in that sense an EPR pair is **maximally entangled**. We conclude by mentioning that we will later see that there always exists a basis such that (1) is perfectly satisfied, and then the *amount* of entanglement is characterized by the $|\alpha_i|^2$'s, and their closeness to the uniform distribution.

3.2 Quantum Teleportation

By **teleportation**, we understand the process of transporting a physical system from one place to another, without actually moving the physical system through the intervening space, merely classical information is communicated. We capture this by the following theorem.

Theorem 3.1. *Let $\mathcal{H}_A = \mathcal{H}_B = \mathcal{H}_E = \mathbb{C}^2$, and let $|\Phi\rangle \in \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_B)$ be an EPR pair. Then, there exists $\mathbf{M} \in \mathcal{M}eas_I(\mathcal{H}_E \otimes \mathcal{H}_A)$ and a family $\{U_i\}_{i\in I}$ of unitaries in $\mathcal{U}(\mathcal{H}_B)$, such that for every $|\varphi\rangle \in \mathcal{S}(\mathcal{H}_E)$ and $i \in I$ there exists $|\psi^i\rangle \in \mathcal{S}(\mathcal{H}_E)$ such that*

$$\big(\mathbb{I}_{EA} \otimes U_i\big)\big(\tfrac{1}{\sqrt{p_i}} M_i \otimes \mathbb{I}_B\big)|\varphi\rangle|\Phi\rangle = |\psi^i\rangle \otimes |\varphi\rangle\,.$$

In other words, Alice can teleport the state $|\varphi\rangle \in \mathcal{S}(\mathcal{H}_E)$ to Bob by measuring EA and sending the measurement outcome i to Bob, and Bob can recover $|\varphi\rangle$ by applying U_i to B (see Figure 3.1).

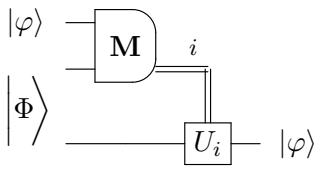


Figure 3.1: Quantum teleportation.

The measurement on Alice's side is given by the **Bell states**

$$|\Phi^+\rangle = \frac{1}{\sqrt{2}}\big(|0\rangle|0\rangle + |1\rangle|1\rangle\big) \qquad\qquad |\Psi^+\rangle = \frac{1}{\sqrt{2}}\big(|0\rangle|1\rangle + |1\rangle|0\rangle\big)$$

$$|\Phi^-\rangle = \frac{1}{\sqrt{2}}\big(|0\rangle|0\rangle - |1\rangle|1\rangle\big) \qquad |\Psi^-\rangle = \frac{1}{\sqrt{2}}\big(|0\rangle|1\rangle - |1\rangle|0\rangle\big)\,,$$

which form a basis of the 2-qubit state space $\mathbb{C}^2 \otimes \mathbb{C}^2$, and the first state, $|\Phi^+\rangle$, is an EPR pair. As can easily be verified, the inverse basis transformation is given by

$$|0\rangle|0\rangle = \frac{1}{\sqrt{2}}(|\Phi^+\rangle + |\Phi^-\rangle) \qquad |1\rangle|0\rangle = \frac{1}{\sqrt{2}}(|\Psi^+\rangle - |\Psi^-\rangle)$$
$$|0\rangle|1\rangle = \frac{1}{\sqrt{2}}(|\Psi^+\rangle + |\Psi^-\rangle) \qquad |1\rangle|1\rangle = \frac{1}{\sqrt{2}}(|\Phi^+\rangle - |\Phi^-\rangle)\,.$$

Proof. $\mathbf{M}$ is the projective rank-1 measurement given by the Bell basis; the unitaries, we will fix later. Consider now an arbitrary qubit state $|\varphi\rangle = \alpha|0\rangle + \beta|1\rangle \in \mathcal{H}_{\mathsf{E}}$. The joint state is then given by the 3-qubit state

$$\begin{aligned}|\Omega\rangle = |\varphi\rangle \otimes |\Phi^+\rangle &= (\alpha|0\rangle + \beta|1\rangle) \otimes \frac{1}{\sqrt{2}}\big(|0\rangle|0\rangle + |1\rangle|1\rangle\big) \\ &= \frac{\alpha}{\sqrt{2}}|0\rangle|0\rangle|0\rangle + \frac{\alpha}{\sqrt{2}}|0\rangle|1\rangle|1\rangle + \frac{\beta}{\sqrt{2}}|1\rangle|0\rangle|0\rangle + \frac{\beta}{\sqrt{2}}|1\rangle|1\rangle|1\rangle\,.\end{aligned}$$

To understand the effect of the Bell measurement, we rewrite the two qubits that Alice controls in the Bell basis, and obtain

$$|\Omega\rangle = \frac{1}{2}|\Phi^+\rangle\big(\alpha|0\rangle + \beta|1\rangle\big) + \frac{1}{2}|\Phi^-\rangle\big(\alpha|0\rangle - \beta|1\rangle\big) + \frac{1}{2}|\Psi^+\rangle\big(\beta|0\rangle + \alpha|1\rangle\big) + \frac{1}{2}|\Psi^-\rangle\big(-\beta|0\rangle + \alpha|1\rangle\big)\,.$$

Hence, depending on the measurement outcome, the state collapses to one of the following four states.

$$|\Phi^+\rangle\big(\alpha|0\rangle + \beta|1\rangle\big), \quad |\Phi^-\rangle\big(\alpha|0\rangle - \beta|1\rangle\big), \quad |\Psi^+\rangle\big(\beta|0\rangle + \alpha|1\rangle\big), \quad \text{or} \quad |\Psi^-\rangle\big(-\beta|0\rangle + \alpha|1\rangle\big)$$

Note the similarity of Bobs qubit to the original state $|\varphi\rangle = \alpha|0\rangle + \beta|1\rangle$; in all cases, the qubit Bob controls can be transformed into the right state by a suitable unitary: in case of the first of the four possible measurement outcomes, it is simply the identity, in case of the second possible outcome, it is the unitary that maps $|0\rangle$ to $|0\rangle$ and $|1\rangle$ to $-|1\rangle$, etc. □

3.3 Violation of Bell's Inequality

An additional peculiar consequence of entanglement is the so-called **nonlocality** of quantum mechanics, sometimes also referred to as a **violation of Bell's inequality**. We use the framework of **nonlocal games** to present it.

We consider two parties, Alice and Bob, and we pose Alice a "question" $x \in \mathcal{X}$ and Bob a "question" $y \in \mathcal{Y}$, and they have to provide respective replies $a \in \mathcal{A}$ and $b \in \mathcal{B}$ by performing *local operations only*, and *without communicating*. If their answers are "good" then they *win*, otherwise they lose. The example we study here is the so-called **CHSH game**, named after Clauser, Horne, Shimony and Holt. Here, Alice and Bob are given *random and independent* bits $x, y \in \mathbb{F}_2$ as "question", and they win if and only if their replies $a, b \in \mathbb{F}_2$ satisfy

$$a + b = x \cdot y\,.$$

What is interesting about such non-local games is that Alice and Bob can have a larger winning probability when they have "quantum capabilities". Formally, the **classical** or **local value** of the CHSH game is defined as

$$v_{\text{CHSH}} := \max \frac{1}{4}\big|\{x, y \in \mathbb{F}_2 \,|\, a(x) + b(y) = x \cdot y\}\big|$$

where the maximum is over all functions $a, b : \mathbb{F}_2 \to \mathbb{F}_2$, and the **quantum value** as

$$v^*_{\text{CHSH}} := \sup \frac{1}{4} \sum_{x,y} \sum_{\substack{a,b \text{ s.t.} \\ a+b=x\cdot y}} \langle\varphi|(E^x_a \otimes F^y_b)|\varphi\rangle$$

where the supremum is over all Hilbert spaces $\mathcal{H}_A, \mathcal{H}_B$, state vectors $|\varphi\rangle \in \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_B)$, and POVM's $\{E^0_0, E^0_1\}, \{E^1_0, E^1_1\} \in \mathcal{POVM}_{\{0,1\}}(\mathcal{H}_A)$ and $\{F^0_0, F^0_1\}, \{F^1_0, F^1_1\} \in \mathcal{POVM}_{\{0,1\}}(\mathcal{H}_B)$.

Since any classical strategy can be "simulated" by means of a quantum strategy, it holds that $v_{\text{CHSH}} \leq v^*_{\text{CHSH}}$. What we will show is that the inequality is strict, i.e., $v_{\text{CHSH}} < v^*_{\text{CHSH}}$.

By a brute-force analysis over all classical strategies, i.e., all functions $a, b : \mathbb{F}_2 \to \mathbb{F}_2$, one can easily check that $v_{\text{CHSH}} = \frac{3}{4}$. As for v^*_{CHSH}, we have the following.

Theorem 3.2. $v^*_{\text{CHSH}} \geq \cos(\frac{\pi}{8})^2 = \frac{1}{2} + \frac{1}{2\sqrt{2}} \approx 0.85$.

Proof. We prove here that $v^*_{\text{CHSH}} \geq \cos(\frac{\pi}{8})^2$; the equality follows from Proposition 3.3 below. We specify a quantum strategy that achieves the claimed value. $\mathcal{H}_A$ and $\mathcal{H}_B$ are qubit systems, and the shared state is taken to be an EPR pair

$$|\varphi\rangle = |\Phi\rangle = \frac{1}{\sqrt{2}}(|0\rangle|0\rangle + |1\rangle|1\rangle) = \frac{1}{\sqrt{2}}(H|0\rangle \otimes H|0\rangle + H|1\rangle \otimes H|1\rangle) \in \mathcal{H}_A \otimes \mathcal{H}_B\,.$$

The POVM's are respectively given by the bases $\{|0\rangle, |1\rangle\}$ and $\{H|0\rangle, H|1\rangle\}$ on Alice's side, i.e. $E^0_0 = |0\rangle\langle 0|$ and $E^0_1 = |1\rangle\langle 1|$ etc., and by $\{B_0|0\rangle, B_0|1\rangle\}$ and $\{B_1|0\rangle, B_1|1\rangle\}$ on Bob's side, where the latter are given by

$$\begin{aligned} B_0|0\rangle &= \cos(\tfrac{\pi}{8})|0\rangle + \sin(\tfrac{\pi}{8})|1\rangle & \qquad B_1|0\rangle &= \cos(\tfrac{\pi}{8})|0\rangle - \sin(\tfrac{\pi}{8})|1\rangle \\ B_0|1\rangle &= -\sin(\tfrac{\pi}{8})|0\rangle + \cos(\tfrac{\pi}{8})|1\rangle & \qquad B_1|1\rangle &= \sin(\tfrac{\pi}{8})|0\rangle + \cos(\tfrac{\pi}{8})|1\rangle \end{aligned}$$

and referred to as **Breidbart bases**. The four bases are illustrated in Figure 3.2.

We will show that $p(win|x,y) := \sum_a \langle\varphi|(E^x_a \otimes F^y_{xy-a})|\varphi\rangle = \cos(\frac{\pi}{8})^2$ for all $x, y \in \mathbb{F}_2$. First, we indeed see that

$$p(win|0,y) = \sum_a \langle\varphi|(E^0_a \otimes F^y_a)|\varphi\rangle = \frac{1}{2}|\langle 0|B_y|0\rangle|^2 + \frac{1}{2}|\langle 1|B_y|1\rangle|^2 = \cos(\tfrac{\pi}{8})^2\,.$$

The case $x = 1$ can be seen equally easily read out from Figure 3.2; the formal computation is slightly more involved. We use that $H|0\rangle$ and $H|0\rangle$ can be written as

$$H|0\rangle = \cos(\tfrac{\pi}{4})|0\rangle + \sin(\tfrac{\pi}{4})|1\rangle \qquad \text{and} \qquad H|1\rangle = \sin(\tfrac{\pi}{4})|0\rangle - \cos(\tfrac{\pi}{4})|1\rangle\,,$$

and we rely on some basic laws of trigonometry. For $x = 1$ and $y = 0$ we then get

$$\begin{aligned} p(win|1{,}0) &= \sum_a \langle\varphi|(E^1_a \otimes F^0_a)|\varphi\rangle = \frac{1}{2}|\langle 0|HB_0H|0\rangle|^2 + \frac{1}{2}|\langle 1|HB_0H|1\rangle|^2 \\ &= \left|\cos(\tfrac{\pi}{4})\cos(\tfrac{\pi}{8}) + \sin(\tfrac{\pi}{4})\sin(\tfrac{\pi}{8})\right|^2 = \left|\cos(\tfrac{\pi}{4} - \tfrac{\pi}{8})\right|^2 = \cos(\tfrac{\pi}{8})^2\,, \end{aligned}$$

and similarly for the final case $x = 1$ and $y = 1$

$$\begin{aligned} p(win|1{,}1) &= \sum_a \langle\varphi|(E^1_a \otimes F^1_{1-a})|\varphi\rangle = \frac{1}{2}|\langle 0|HB_1H|1\rangle|^2 + \frac{1}{2}|\langle 1|HB_1H|0\rangle|^2 \\ &= \left|\sin(\tfrac{\pi}{4})\cos(\tfrac{\pi}{8}) + \cos(\tfrac{\pi}{4})\sin(\tfrac{\pi}{8})\right|^2 = \left|\sin(\tfrac{\pi}{4} + \tfrac{\pi}{8})\right|^2 = \cos(\tfrac{\pi}{8})^2\,, \end{aligned}$$

This concludes the proof. □

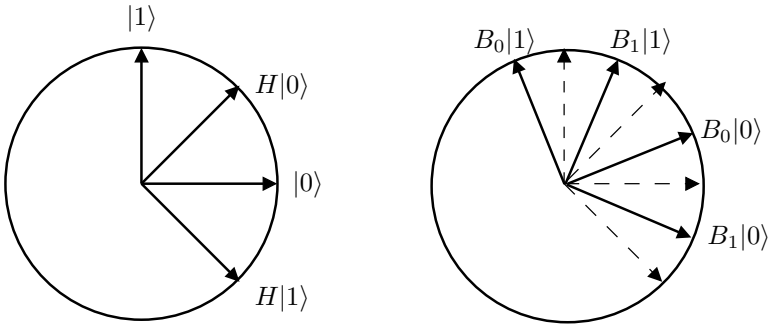


Figure 3.2: Quantum strategy for CHSH.

We now consider the natural extension of the CHSH game to an arbitrary prime-field $\mathbb{F}_p$, and we define the quantum value $v^*_{\mathrm{CHSH}(p)}$ accordingly. Then, the following upper bound holds.

Proposition 3.3. *$v^*_{\mathrm{CHSH}(p)} \leq \frac{1}{p} + \frac{p-1}{p\sqrt{p}}$ for any prime $p \in \mathbb{N}$.*

Proof. Consider an arbitrary quantum strategy and set $P_{AB|XY}(a,b|x,y) := \langle\varphi|(E_a^x \otimes F_b^y)|\varphi\rangle$. By Naimark dilation (Theorem 2.3), we may assume that the E_a^x's and F_b^y's are projections. Let $\omega := e^{2\pi i/p} \in \mathbb{C}$, or any other primitive p-th root of unity. We consider the **Fourier transform** of $P_{AB|XY}(a,b|x,y)$

$$\begin{aligned}\hat{P}_{AB|XY}(k,\ell|x,y) &= \frac{1}{p}\sum_{a,b}\omega^{ka}\omega^{\ell b}P_{AB|XY}(a,b|x,y)\\ &= \frac{1}{p}\sum_{a,b}\omega^{ka}\omega^{\ell b}\langle\varphi|(E_a^x\otimes F_b^y)|\varphi\rangle = \frac{1}{p}\langle\varphi|U_x^k\otimes V_y^\ell|\varphi\rangle = \frac{1}{p}\langle u_x^k|v_y^\ell\rangle\,,\end{aligned}$$

where $U_x^k := \sum_a \omega^{ka}E_a^x$ and $V_y^\ell := \sum_b \omega^{\ell b}F_b^y$, and $|u_x^k\rangle := (U_x^k\otimes\mathbb{I})^\dagger|\varphi\rangle$ and $|v_y^\ell\rangle := (\mathbb{I}\otimes V_y^\ell)|\varphi\rangle$. What is important here is that U_x^k and V_y^ℓ are *unitary*, and thus $|u_x^k\rangle$ and $|v_y^\ell\rangle$ have norm 1. That U_x^k is unitary is easily verified given that for every x the E_a^x's are pairwise mutually orthogonal projections that add up to $\mathbb{I}$, and similarly for V_y^ℓ.

From the observation that $\sum_k \omega^{kd}$ vanishes for $d \neq 0$ and equals p for $d = 0$, it follows that

$$P_{AB|XY}(a,b|x,y) = \frac{1}{p}\sum_{k,\ell}\omega^{-ka}\omega^{-\ell b}\hat{P}_{AB|XY}(k,\ell|x,y) = \frac{1}{p^2}\sum_{k,\ell}\omega^{-ka}\omega^{-\ell b}\langle u_x^k|v_y^\ell\rangle$$

and thus

$$\begin{aligned}\sum_{\substack{a,b\text{ s.t.}\\ a+b=xy}} P_{AB|XY}(a,b|x,y) &= \frac{1}{p^2}\sum_{\substack{a,b\text{ s.t.}\\ a+b=xy}}\sum_{k,\ell}\omega^{-ka}\omega^{-\ell b}\langle u_x^k|v_y^\ell\rangle = \frac{1}{p^2}\sum_a\sum_{k,\ell}\omega^{-ka}\omega^{-\ell(xy-a)}\langle u_x^k|v_y^\ell\rangle\\ &= \frac{1}{p^2}\sum_a\sum_{k,\ell}\omega^{-(k-\ell)a}\omega^{-\ell xy}\langle u_x^k|v_y^\ell\rangle = \frac{1}{p}\sum_k\omega^{-kxy}\langle u_x^k|v_y^k\rangle\,.\end{aligned}$$

We conclude that

$$\begin{aligned}\frac{1}{p^2}\sum_{x,y}\sum_{\substack{a,b\text{ s.t.}\\ a+b=xy}} P_{AB|XY}(a,b|x,y) &= \frac{1}{p^3}\sum_{x,y}\sum_k\omega^{-kxy}\langle u_x^k|v_y^k\rangle\\ &= \frac{1}{p^3}\sum_{x,y}\langle u_x^0|v_y^0\rangle + \frac{1}{p^3}\sum_{k\neq 0}\sum_{x,y}\omega^{-kxy}\langle u_x^k|v_y^k\rangle\,,\end{aligned}$$

and the bound $v^*_{\mathrm{CHSH}(p)} \leq \frac{1}{p} + \frac{p-1}{p\sqrt{p}}$ follows from Cauchy-Schwarz and the lemma below (noting that ω^k is a primitive p-th root of unity for any $k \neq 0$). □

Lemma 3.4. *Let p be a prime, let $\omega \in \mathbb{C}$ be a primitive p-th root of unity, and let $\{|u_x\rangle\}_{x\in\mathbb{F}_p}$ and $\{|v_y\rangle\}_{y\in\mathbb{F}_p}$ be two families of vectors in $\mathcal{S}(\mathcal{H})$. Then*

$$\left|\sum_{x,y} \omega^{-xy} \langle u_x|v_y\rangle\right| \leq p\sqrt{p}\,.$$

Proof. We set $|u\rangle := \sum_x |x\rangle|u_x\rangle$ and $|v\rangle := \sum_y |y\rangle|v_y\rangle$, where $\{|0\rangle, \ldots, |p-1\rangle\}$ is a basis of $\mathbb{C}^p$. Note that $\||u\rangle\|^2 = \sum_x \langle u_x|u_x\rangle = p$, and the same for $\||v\rangle\|^2$. Furthermore, we consider

$$|\check{v}_x\rangle := \frac{1}{\sqrt{p}} \sum_y \omega^{-xy} |v_y\rangle \in \mathcal{H}$$

and, like above, set $|\check{v}\rangle := \sum_x |x\rangle|\check{v}_x\rangle$. Note that

$$\||\check{v}\rangle\|^2 = \sum_x \langle \check{v}_x|\check{v}_x\rangle = \frac{1}{p} \sum_{x,y,y'} \omega^{x(y-y')} \langle v_y|v_{y'}\rangle = \sum_y \langle v_y|v_y\rangle = \||v\rangle\|^2$$

(Parseval's identity). Using Cauchy-Schwarz, we now see that

$$\left|\sum_{x,y} \omega^{-xy} \langle u_x|v_y\rangle\right| = \sqrt{p} \left|\sum_x \langle u_x|\check{v}_x\rangle\right| = \sqrt{p}\,\langle u|\check{v}\rangle \leq \sqrt{p}\,\||u\rangle\|\||\check{v}\rangle\| = \sqrt{p}\,\||u\rangle\|\||v\rangle\| = p\sqrt{p}\,.$$

□

Part II

Quantum Computing

Chapter 4

Foundations of Quantum Computing

Quantum computation investigates the computational power of hypothetical computing devices that make use of quantum-mechanical properties, as introduced and discussed in previous chapters. An important objective is to find quantum algorithms that are significantly faster than any classical algorithm solving the same problem. The field started in the early 1980s with suggestions for analog quantum computers by Yuri Manin, Richard Feynman, and Paul Benioff, and reached more digital ground when in 1985 David Deutsch defined the universal quantum Turing machine. The following years saw only sparse activity, notably the development of the first algorithms by Deutsch and Jozsa and by Simon, and the development of quantum complexity theory by Bernstein and Vazirani. Interest in the field then increased tremendously after Peter Shor's 1994 discovery of his famous quantum algorithms for factoring large integers and for computing discrete logarithms.

In this chapter, we introduce some of the early quantum algorithms, and we cover the theoretical foundations by discussing the quantum circuit model of computation. We end the chapter with Grover's algorithm for unstructured search, a quantum algorithm that is maybe not as impressive in terms of speed-up (like Shor's algorithms) but is important due to the generality of the computational problem it solves.

In this chapter, if not specified otherwise, we restrict $\mathcal{H}$ to be $\mathcal{H} = \mathbb{C}^2$, the state space of a qubit. For any positive integer n, $\mathcal{H}^{\otimes n}$ stands for the n-fold tensor product $\mathcal{H} \otimes \ldots \otimes \mathcal{H}$ of $\mathcal{H}$ with itself. Similarly, for $U \in \mathcal{U}(\mathcal{H})$, $U^{\otimes n}$ is the n-fold tensor product $U \otimes \ldots \otimes U \in \mathcal{U}(\mathcal{H}^{\otimes n})$. Throughout, we consider the computational basis $\{|0\rangle, |1\rangle\}$ of $\mathcal{H}$ as well as the computational basis $\{|x\rangle\}_{x\in\{0,1\}^n}$ of $\mathcal{H}^{\otimes n}$, where $|x\rangle = |x_1, \ldots, x_n\rangle = |x_1\rangle \cdots |x_n\rangle$ for $x = (x_1, \ldots, x_n) \in \{0,1\}^n$.

4.1 Warm-up: Deutsch's Algorithm

Consider a binary function $f : \{0,1\} \to \{0,1\}$. We imagine a situation where f is not given to us by its function table, but, say, in the form of a very complicated and extremely inefficient algorithm. As such, we can learn $f(x)$ for any $x \in \{0,1\}$ by *computing* $f(x)$ using the algorithm, but there is no "shortcut". One typically speaks of **black-box access** then: the only way to learn (anything about) the function value of any input x is by means of making a **query** to an "oracle", which then provides the correct function value $f(x)$.

The task here now is to find out if $f(0) = f(1)$ or not. Obviously, this can be done by computing f *twice*, i.e., by making two queries — one for input 0 and one for input 1. The question is whether one can do better. It is intuitively quite clear, and not too hard to prove once rigorously formalized, that it is impossible to do any better with a classical algorithm: any classical algorithm with black-box access to f that only makes *one* query to f cannot

predict whether $f(0) = f(1)$ or not with probability larger than $\frac{1}{2}$, when f is chosen uniformly at random from all functions $\{0,1\} \to \{0,1\}$. On the other hand, somewhat surprisingly, a *quantum* algorithm with black-box *quantum access* to f can do better. The latter means that the algorithm can make queries to the unitary $U_f \in \mathcal{U}(\mathcal{H} \otimes \mathcal{H})$ given by $U_f|x\rangle|y\rangle = |x\rangle|y \oplus f(x)\rangle$ for $x, y \in \{0,1\}$, where $\oplus$ is the binary XOR, i.e., addition mod 2.

Proposition 4.1. *There exists a quantum algorithm so that when given black-box access to U_f it makes a single query to U_f and outputs the bit $f(0) \oplus f(1)$ with certainty.*

Note that we do not have a formal notion of a "quantum algorithm" yet, but it should be clear that what we outline in the proof counts as one. It is called **Deutsch's algorithm**, named after David Deutsch. Before describing the algorithm, let us first look into the naive approach. Given that the desired quantum algorithm has access to U_f, which means that he can apply f to a *superposition* of inputs, that seems to be the way to go: form the superposition $(|0\rangle + |1\rangle)/\sqrt{2}$ and apply U_f to obtain

$$U_f \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)|0\rangle = \frac{1}{\sqrt{2}}|0\rangle|f(0)\rangle + \frac{1}{\sqrt{2}}|1\rangle|f(1)\rangle\,.$$

This indeed gives us a state that depends on *both*, $f(0)$ and $f(1)$. However, it is unclear how to now *extract information* on both, respectively on $f(0) \oplus f(1)$. If we measure the first qubit, then the second collapses to $|f(0)\rangle$ or $|f(1)\rangle$, and all information on the other is lost, and so we would still need another call to f. Also, if we measure the second qubit it is easy to see that it does not provide any information on $f(0) \oplus f(1)$.

Proof. The algorithm starts with the 2-qubit state

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

and applies U_f. This yields

$$\begin{aligned}
&\frac{1}{2}|0\rangle \otimes \big(|0 \oplus f(0)\rangle - |1 \oplus f(0)\rangle\big) + \frac{1}{2}|1\rangle \otimes \big(|0 \oplus f(1)\rangle - |1 \oplus f(1)\rangle\big) \\
&\qquad = \frac{1}{2}(-1)^{f(0)}\,|0\rangle \otimes \big(|0\rangle - |1\rangle\big) + \frac{1}{2}(-1)^{f(1)}\,|1\rangle \otimes \big(|0\rangle - |1\rangle\big) \\
&\qquad = \frac{1}{\sqrt{2}}\Big(|0\rangle + (-1)^{f(0)\oplus f(1)}|1\rangle\Big) \otimes \frac{1}{\sqrt{2}}(-1)^{f(0)}\big(|0\rangle - |1\rangle\big)\,.
\end{aligned}$$

which is still a product state, and it has the predicate we are looking for encoded into the phase of the first qubit. Concretely, the first qubit is in state $|+\rangle$ if $f(0) = f(1)$, and it is in state $|-\rangle$ if $f(0) \neq f(1)$. Hence, by now measuring the first qubit in the Hadamard basis, or, equivalently, by applying the Hadamard operator and measuring in the computational basis, we obtain the correct result. □

Clearly, Deutsch's algorithm does not seem very relevant nor impressive from a practical perspective. Still, it nicely shows, in a simple way, how it is still possible to exploit the possibility of applying a function f to a quantum superposition of inputs, despite the observation that the naive approach does not give you anything. The idea here is to bring the function values from the basis vectors into the amplitudes, so that one gets *constructive* or *distructive interference*, depending on the function values. Here, this is achieved by applying U_f to $|x\rangle|-\rangle$ for $x \in \{0,1\}$, rather than to $|x\rangle|0\rangle$, so that we get

$$U_f|x\rangle|-\rangle = \frac{1}{\sqrt{2}}\big(|x\rangle|f(x)\rangle - |x\rangle|1 \oplus f(x)\rangle\big) = (-1)^{f(x)}|x\rangle|-\rangle\,.$$

This technique is typically referred to as **phase kickback**. Put differently, we use the fact that U_f has eigenvector $|x\rangle|-\rangle$ with eigenvalue $(-1)^{f(x)}$. Effectively, this gives access to a unitary $V_f \in \mathcal{U}(H)$ with $V_f|x\rangle = (-1)^{f(x)}|x\rangle$, and we now easily see that if $f(0) = f(1)$ then $V_f|+\rangle \equiv |+\rangle$ and if $f(0) \neq f(1)$ then $V_f|+\rangle \equiv |-\rangle$, and so we can distinguish the two cases with one call to V_f.

4.2 More Examples: Deutsch-Jozsa and Bernstein-Vazirani

Deutsch's algorithm is a special case of the **Deutsch-Jozsa algorithm**. The latter considers a function $f : \{0,1\}^n \to \{0,1\}$ with the promise that f is either constant or balanced, where the latter means that $|\{x \mid f(x) = 0\} = 2^{n-1}$, and the goal is to find out which of the two is the case. Classically, this requires $2^{n-1} + 1$ queries to f in the worst case, and k queries to get the right answer except with probability 2^{-k+1}. With a quantum algorithm, one query suffices.

Proposition 4.2. *For any positive integer $n \in \mathbb{N}$ there exists a quantum algorithm so that when given black-box access to $U_f \in \mathcal{U}(\mathcal{H}^{\otimes n} \otimes \mathcal{H})$ it makes a single query to U_f and predicts with certainty whether $f : \{0,1\}^n \to \{0,1\}$ is constant or balanced, given that it is one of the two.*

For the analysis of the algorithm, it will be convenient to have the following lemma at hand; we leave its proof as an exercise. $H^{\otimes n}$ is sometimes also called **Walsh-Hadamard transform**.

Lemma 4.3. *For any $x = (x_1, \ldots, x_n) \in \{0,1\}^n$,*

$$H^{\otimes n}|x\rangle = H|x_1\rangle \otimes \ldots \otimes H|x_n\rangle = \frac{1}{2^{n/2}} \sum_{y \in \{0,1\}^n} (-1)^{x \cdot y}|y\rangle$$

where $x \cdot y = x_1 y_1 \oplus \ldots \oplus x_n y_n \in \{0,1\}$.

Proof (of Proposition 4.2). The algorithm follows closely Deutsch's algorithm. It starts off with the $(n+1)$-qubit state $|+\rangle^{\otimes n} \otimes |-\rangle$, which equals

$$H^{\otimes n}|0,\ldots,0\rangle \otimes H|1\rangle = \frac{1}{2^{n/2}} \sum_{x \in \{0,1\}^n} |x\rangle \otimes \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle)$$

by Lemma 4.3, and it applies U_f. This yields

$$\frac{1}{2^{n/2}} \sum_x |x\rangle \otimes \frac{1}{\sqrt{2}}\big(|f(x)\rangle - |1 \oplus f(x)\rangle\big) = \frac{1}{2^{n/2}} \sum_x (-1)^{f(x)}|x\rangle \otimes \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle)\,.$$

Again, we ignore the last qubit, which is in "product form", and apply $H^{\otimes n}$ to the first n qubits. This results in the n-qubit state

$$\frac{1}{2^{n/2}} \sum_x (-1)^{f(x)} \frac{1}{2^{n/2}} \sum_y (-1)^{x \cdot y}|y\rangle = \frac{1}{2^n} \sum_y \left(\sum_x (-1)^{f(x)}(-1)^{x \cdot y} \right)|y\rangle\,,$$

using again Lemma 4.3. Measuring this state in the computational basis $\{|y\rangle\}_{y \in \{0,1\}^n}$, we see that measurement outcome $y = (0,\ldots,0)$ is observed with probability

$$p_{(0,\ldots,0)} = \left| \frac{1}{2^n} \sum_x (-1)^{f(x)} \right|^2,$$

which is 1 if f is constant and 0 if f is balanced. □

The same algorithm can also be used for finding $s \in \{0,1\}^n$ when given black-box quantum access to the function $f_s : \{0,1\}^n \to \{0,1\}$, $x \mapsto s \cdot x$. In this context, it is then referred to as the **Bernstein-Vazirani algorithm**.

Proposition 4.4. *For any $n \in \mathbb{N}$ there is a quantum algorithm so that when given black-box access to $U_{f_s} \in \mathcal{U}(\mathcal{H}^{\otimes n} \otimes \mathcal{H})$ it makes a single query to U_{f_s} and outputs s with certainty.*

Proof. We consider the n-qubit state obtained by means of the Deutsch-Jozsa algorithm:

$$\frac{1}{2^n} \sum_{x,y} (-1)^{f_s(x)} (-1)^{x \cdot y} |y\rangle = \frac{1}{2^n} \sum_{y} \left(\sum_{x} (-1)^{x \cdot (s \oplus y)} \right) |y\rangle \,,$$

and we observe that the measurement outcome s is observed with probability 1. □

4.3 Quantum Algorithms and Complexity

We want to formalize the notion of a *quantum algorithm* and of the *complexity* of such an algorithm. Indeed, what is exciting about quantum computation is that, for certain computational problems, it allows for algorithms with (much) better complexity compared to classical models of computation. In the above examples, this was demonstrated for the notion of **query complexity**, i.e., when one considers algorithms that make black-box queries to some (partly) unknown "oracle" (typically a function), and we count the number of queries necessary to perform the desired computation. What is nice about this notion of complexity is that it allows for provable lower bounds, and thus for provable separation results between classical and quantum computation.

But maybe more relevant from a practical perspective is the notion of **computational complexity**, which counts the number of "elementary steps" that the algorithm applies to the input. Here, an "elementary step" would be an "elementary unitary operation", which in this context is then referred to as a **gate**.

In order to formalize the above, let $\mathcal{G}$ be a non-empty set of such gates, i.e., of unitary operators, such that for every $G \in \mathcal{G}$ we have that $G \in \mathcal{U}(\mathcal{H}^{\otimes k})$ for some $k \leq n$, called the **arity** of G. For these gates to be "elementary" we will later require the arity k for every $G \in \mathcal{G}$ to be small, like at most 2.

In line with previous observations, a gate $G \in \mathcal{U}(\mathcal{H}^{\otimes k})$ with arity k can act on a n-qubit state (vector); but it then needs to be specified on which k of the n qubits, and which component of G acts on which of the designated k qubits. By default, we label the n qubits by "1","2",...,"n", and then write, say, $G_{7,2,8,5}$ to specify that the first component of $G \in \mathcal{U}(\mathcal{H}^{\otimes 4})$ acts on qubit "7", the second on qubit "2", etc. Similarly, $CNOT_{4,1}$ then refers to the CNOT gate that is controlled by qubit "4" and has target qubit "1". However, we will make little use of this notation; instead, we will mainly use pictorial descriptions (see e.g. Figure 4.1).

We can now define the following model of quantum computation, which is a notion of a quantum algorithm with *quantum input* and *quantum output*.

Definition 4.1. *An n-qubit* **quantum circuit** *with gate set $\mathcal{G}$ consists of a finite sequence $U_1, \ldots, U_t \in \mathcal{U}(\mathcal{H}^{\otimes n})$, where, for every $i \in \{1, \ldots, t\}$, U_i is of the form $U_i = G_W$ with $G \in \mathcal{G}$ and $W = (w_1, \ldots, w_k)$ a sequence of k distinct numbers in $\{1, \ldots, k\}$, where k is the arity of G. The* **computational complexity** *of such a quantum circuit is given by t.*

Remark 4.1. Sometimes, we consider a relaxation of the the above definition where $\mathcal{H}^{\otimes n}$ is replaced by the n-fold tensor product $\mathcal{H}_1 \otimes \cdots \otimes \mathcal{H}_n$ of possibly non-qubit state spaces.

A quantum circuit can be nicely depictured with "wires" and "gates" that act on the "wires", as we have already been doing, e.g. in Section 2.5. Figure 4.1 shows such a pictorial representation of an example quantum circuit with gate set $\mathcal{G} = \{H, CNOT\}$.

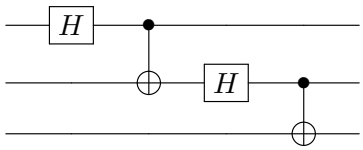


Figure 4.1: Pictorial representation of the quantum circuit $H_1, CNOT_{1,2}, H_2, CNOT_{2,3}$.

Definition 4.2. *An n-qubit quantum circuit $U_1, \ldots, U_t$* **computes** *a unitary $U \in \mathcal{U}(\mathcal{H}^{\otimes n})$ if $U = U_t \cdots U_1$; it ε-***approximately computes** *U for $0 \leq \varepsilon \leq 1$ if $\|U - U_t \cdots U_1\| \leq \varepsilon$.*
More generally, for $n' \geq n$, an n'-qubit quantum circuit $U_1, \ldots, U_t$ is said to **(approximately) compute** *an n-qubit unitary $U \in \mathcal{U}(\mathcal{H}^{\otimes n})$ if $U_t \cdots U_1 |\varphi\rangle |0\rangle^{\otimes(n'-n)}$ equals (or approximates) $U|\varphi\rangle \otimes |0\rangle^{\otimes(n'-n)}$ for any $|\varphi\rangle \in \mathcal{S}(\mathcal{H}^n)$.*

The choice of the norm for measuring the distance between the unitaries is not so important; for concreteness, we take the operator norm[1] here. Similarly for the second part of the definition, where we take the norm induced by the inner product, but the fidelity is also a common choice.

Remark 4.2. Repeating Remark 2.1, we stress that the convention is to apply such a quantum circuit as depictured in Figure 4.1 *from the left to the right*, i.e., first H is applied and then *CNOT* etc., whereas the composition $U = U_t \cdots U_1$ is applied to a state $|\varphi\rangle$ *from right to left.*

Looking back to the decomposition of (multi-)control unitaries, the right hand side of Figure 2.3 can now be formally understood as (a pictorial representation of) a quantum circuit with gate set $\mathcal{G} = \{V, V^\dagger, CNOT\}$ that computes $C^2(U)$, and the right hand side of Figure 2.4 as a quantum circuit, with $\mathcal{G}$ consisting of *CNOT* and the Toffoli gate, that computes $C^n(U)$.

Not so surprising, quantum computing is at least as powerful than classical computing; in particular, once everything is formalized, it is not hard to see that the following holds.

Theorem 4.5. *For any set $\mathcal{G}$ of logic gates, like $\{\wedge, \neg\}$, let $f : \{0,1\}^{m_{in}} \to \{0,1\}^{m_{out}}$ be a function that can be classically computed by a classical circuit C with c gates of fan-out at most ℓ. Then, there exists an n-qubit quantum circuit that computes U_f with gate set $\{U_g \,|\, g \in \mathcal{G}\}$ and computational complexity t, where $t \leq 2c\ell$ and $n \leq m_{in} + m_{out} + c\ell$.*

The factor 2 blow-up in complexity comes from the fact that one has to undo all gates that produce an intermediary (qu)bit, in order to revert the state back to $|0\rangle$.

We will mainly be interested in quantum algorithms with *classical input* and *classical output.* Such a notion can be easily obtained by modifying Definition 4.2 as follows.

Definition 4.3. *For any function $f : \{0,1\}^{m_{in}} \to \{0,1\}^{m_{out}}$ with $0 < m_{in}, m_{out} \leq n$, we say that an n-qubit quantum circuit ε-***approximately computes** *f if*

$$\left\| \left(|f(x)\rangle\langle f(x)| \otimes \mathbb{I} \right) U_t \cdots U_1 |x, 0\rangle \right\|^2 \geq 1 - \varepsilon$$

for all $x \in \{0,1\}^{m_{in}}$, where we understand that $|0\rangle \in \mathcal{S}\big(\mathcal{H}^{\otimes(n-m_{in})}\big)$ and $\mathbb{I} \in \mathcal{U}\big(\mathcal{H}^{\otimes(n-m_{out})}\big)$.

[1] $\|A\| = \max_{|\varphi\rangle} \|A|\varphi\rangle\|$ where the max is over all$|\varphi\rangle \in \mathcal{S}(\mathcal{H}^{\otimes n})$.

In other words, the algorithm proceeds by encoding the input into a quantum state, appending an ancilla register, running the quantum circuit, and then measuring (part of) the resulting state. The approximation parameter ε captures the probability of an incorrect outcome.

Sometimes, it will also be convenient to allow some classical "post-processing" of the measurement result obtained after the application of the quantum algorithm; again, this is without loss of generality since, by Theorem 4.5, such a classical "post-processing" could be incorporated into the quantum circuit. Note that as soon as $\varepsilon < 1/2$, one can amplify the success probability by repeating the algorithm.

Finally, we point out that we can easily extend the above to a notion of quantum algorithm *with black-box access* to a non-specified unitary $O \in \mathcal{U}(\mathcal{H}^{\otimes k})$ with a given arity k: we simply extend the set of gates $\mathcal{G}$ to $\mathcal{G} \cup \{O\}$, meaning that the U_i's may also be instructions to apply O to k of the qubits. The **query complexity** is then defined to be the number of i's for which U_i is an instruction to apply O. Unless specified differently, such a quantum algorithm with black-box access acts on the fixed input state $|0\rangle \in \mathcal{S}(\mathcal{H}^{\otimes n})$, and the classical output is obtained by measuring (a specified subset of) the resulting qubits. Once O is instantiated with a specific unitary, we can then make statements about the statistics of the measurement outcome. This finally puts the statements of Propositions 4.1, 4.2 and 4.4 on firm theoretical grounds. See Figure 4.2 below for the quantum circuit for Deutsch's algorithm, where we have "stacked" the two H gates that can be applied in parallel, and the last gate on the upper wire is a measurement in the computational basis.

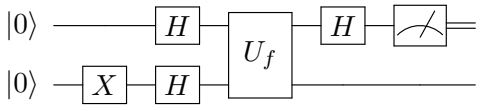


Figure 4.2: Quantum circuit for Deutsch's algorithm.

4.4 Universal Gate Sets

For the notion of quantum circuits to be *complete* as a model of computation, we need a gate set $\mathcal{G}$ that, in principle, enables to compute any unitary.

Definition 4.4. *A (possibly infinite) set of gates $\mathcal{G}$ is called* **perfectly universal** *if for any $n \in \mathbb{N}$ and any unitary $U \in \mathcal{U}(\mathcal{H}^{\otimes n})$ there exists an n-qubit quantum circuit with gate set $\mathcal{G}$ that computes $e^{i\alpha}U$ for some $\alpha \in \mathbb{R}$.*[2]
$\mathcal{G}$ is called **approximately universal** *if for any $n \in \mathbb{N}$, $\varepsilon > 0$ and $U \in \mathcal{U}(\mathcal{H}^{\otimes n})$ there exists an n-qubit quantum circuit with gate set $\mathcal{G}$ that ε-approximately computes $e^{i\alpha}U$ for some $\alpha \in \mathbb{R}$*

The freedom in the phase is motivated by the fact that such a global phase has no noticable effect. As a first step towards obtaining a universal gate set, we show that any unitary can be decomposed into *two-level* unitaries, defined as follows.

Definition 4.5. *For an arbitrary Hilbert space $\mathcal{H}$ with orthonormal basis $\{|i\rangle\}_{i\in I}$, a unitary $U \in \mathcal{U}(\mathcal{H})$ is said to be* **two-level** *(w.r.t. $\{|i\rangle\}_{i\in I}$) if $U|i\rangle = |i\rangle$ for* all but two *choices of $i \in I$.*

In other words, U acts non-trivially only on (at most) two basis vectors $|k\rangle$ and $|\ell\rangle$. It is then easy to see that $U|k\rangle = u_{11}|k\rangle + u_{21}|\ell\rangle$ and $U|\ell\rangle = u_{12}|k\rangle + u_{22}|\ell\rangle$, where $\tilde{U} := \left[\begin{smallmatrix} u_{11} & u_{12} \\ u_{21} & u_{22} \end{smallmatrix}\right] \in \mathcal{U}(\mathbb{C}^2)$, and we write $U = \mathbb{I} \oplus \tilde{U}$. It is also easy to see that $U^\dagger$ is two-level if U is, namely $U^\dagger = \mathbb{I} \oplus \tilde{U}^\dagger$.

[2] In the literature, the terminology *universal* is somewhat ambiguously used for different variations, like for *approximately* or *perfectly* universal, or when allowing the quantum circuit to be over more than n qubits.

Proposition 4.6. *For an arbitrary Hilbert space $\mathcal{H}$ and for any $U \in \mathcal{U}(\mathcal{H})$, there exists two-level unitaries $U_1, \ldots, U_N \in \mathcal{U}(\mathcal{H})$ (w.r.t. any fixed basis) so that $U = U_1 \cdots U_N$.*

Proof. If $U = \mathbb{I}$ then there is nothing to prove, so we consider $U \neq \mathbb{I}$. Thus, there exists a maximal k (for any fixed ordering) so that $x := \langle k|U|k\rangle \neq 1$. We consider now two cases.

Case 1: $\langle k|U|\ell\rangle = 0$ for all $\ell \neq k$. Here, we note that

$$1 = \langle k|k\rangle = \langle k|UU^\dagger|k\rangle = \sum_\ell \langle k|U|\ell\rangle\langle\ell|U^\dagger|k\rangle = \sum_\ell |\langle k|U|\ell\rangle|^2 = |\langle k|U|k\rangle|^2 = |x|^2\,,$$

and then consider the two-level (actually one-level) unitary V that acts non-trivially only on $|k\rangle$, and does so as $V|k\rangle = \bar{x}|k\rangle$. We then set $U' := UV$ and observe that $\langle k|U'|k\rangle = \langle k|U|k\rangle\bar{x} = x\bar{x} = 1$, while still $\langle k'|U'|k'\rangle = \langle k'|U|k'\rangle = 1$ for all $k' > k$.

Case 2: $\langle k|U|\ell\rangle \neq 0$ for some $\ell \neq k$. Let ℓ maximal and $y := \langle k|U|\ell\rangle$. Here,

$$1 = \langle \ell|\ell\rangle = \langle \ell|U^\dagger U|\ell\rangle = \sum_m \langle \ell|U^\dagger|m\rangle\langle m|U|\ell\rangle = \sum_m \langle m|U|\ell\rangle|^2 > |\langle \ell|U|\ell\rangle|^2$$

and thus $\ell < k$ by the maximality of k, and we then consider the two-level unitary V, called **Givens rotation**, that acts only on $|k\rangle$ and $|\ell\rangle$, and does so as

$$V|k\rangle = \bar{u}|k\rangle + \bar{v}|\ell\rangle \qquad \text{and} \qquad V|\ell\rangle = v|k\rangle - u|\ell\rangle$$

where

$$u = \frac{x}{\sqrt{|x|^2 + |y|^2}} \quad \text{and} \quad v = \frac{y}{\sqrt{|x|^2 + |y|^2}}\,.$$

It is easy to verify that V is unitary. Furthermore, setting $U' := UV$, we observe that

$$\langle k|U'|\ell\rangle = v\langle k|U|k\rangle - u\langle k|U|\ell\rangle = vx - uy = 0'\,,$$

while still $\langle k|U'|\ell'\rangle = 0$ for all $\ell' > \ell$ with $\ell' \neq k$, and still $\langle k'|U'|k'\rangle = 1$ for all $k' > k$.

Thus, in either case, by a recursive application of the above to U', we obtain a finite sequence $V_1, \ldots, V_N$ of two-level unitaries so that $UV_1 \cdots V_N = I$. By inverting the two-level unitaries, we obtain the claimed result. □

For the remainder, we again fix $\mathcal{H}$ to be $\mathcal{H} = \mathbb{C}^2$. The following shows that all single-qubit gates together with *CNOT* form a universal set of gates.

Theorem 4.7. *The gate set $\mathcal{G} = \{CNOT\} \cup \mathcal{U}(\mathcal{H})$ is perfectly universal.*

Proof. By Proposition 4.6 above, it is sufficient to show that any two-level unitary $U = \mathbb{I} \oplus \tilde{U}$ (w.r.t. the computational basis) can be computed with $\mathcal{G}$.[3] For this, we first consider the special case where the two basis vectors $|k\rangle$ and $|\ell\rangle$ on which U acts nontrivially are such that $k \in \{0,1\}^n$ and $\ell \in \{0,1\}^n$ differ in only one bit, say, $|k\rangle = |k_1\rangle \cdots |k_{n-1}\rangle|0\rangle$ and $|\ell\rangle = |k_1\rangle \cdots |k_{n-1}\rangle|1\rangle$.[4] In this case, we consider the multi-control unitary $C = C^{n-1}[k_1, \ldots, k_{n-1}](\tilde{U})$ that is controlled by the first $n-1$ qubits to be in state $|k_1, \ldots, k_{n-1}\rangle$ and acts on the last qubit as $\tilde{U}$, and we see that

$$C|k\rangle = |k_1, \ldots, k_{n-1}\rangle \otimes \tilde{U}|0\rangle = |k_1, \ldots, k_{n-1}\rangle \otimes (u_{11}|0\rangle + u_{21}|1\rangle) = u_{11}|k\rangle + u_{21}|\ell\rangle = U|k\rangle$$

[3] Note, however, that the computational complexity of computing general unitaries by means of two-level unitaries is quadratic in $d = 2^n$, and thus exponential in n. Thus, for *efficient* quantum computation, we need more clever ways to compute the unitaries of interest.

[4] That it is the last bit here makes the writing easier but is not crucial to the argument.

and similarly $C|\ell\rangle = U|\ell\rangle$, while $C|i\rangle = |i\rangle$ for any $i \neq k, \ell$. Thus, $C = U$, and by Corollay 2.9 such a multi-control unitary can be computed with $\mathcal{G}$.

In the more typical case where k and ℓ differ in more than one bit, we argue by induction on the number of bits they differ. Choose $k' \in \{0,1\}^n$ such that k' and k differ in one bit, and k' and ℓ in one bit *less* than k and ℓ. Consider the two-level Hermitian unitary V that maps $V|k\rangle = |k'\rangle$ and $V|k'\rangle = |k\rangle$. Then, $VUV|i\rangle = VU|i\rangle = V|i\rangle = |i\rangle$ for any $i \notin \{k, k', \ell\}$, but also $VUV|k\rangle = VU|k'\rangle = V|k'\rangle = |k\rangle$. Thus, the unitary $U' := VUV$ acts non-trivial only on $|k'\rangle$ and $|\ell\rangle$, and so we can apply the induction hypothesis to V and U'. Given that $U = VU'V$, this then proves the claim. □

In combination with Theorem 1.4, we immediately get the following "smaller" gate set.

Corollary 4.8. *The gate set* $\mathcal{G} = \{CNOT\} \cup \bigcup_{0 \leq \theta < 4\pi} \{R_Y(\theta), R_Z(\theta)\}$ *is perfectly universal.*

We conclude the discussion of universal gate sets with the following two fundamental results, which we state here without (full) proofs. The first result shows that we can replace the uncountable set of 1-qubit gates in the universal gate set $\mathcal{G}$ considered above by two particular 1-qubit gates and still get an *approximate* universal gate set, and the second result shows that these two single 1-qubit gates approximate any 1-qubit gate with low computational complexity.

Theorem 4.9. *For any* $n \in \mathbb{N}$, $\varepsilon > 0$ *and* $U \in \mathcal{U}(\mathcal{H}^{\otimes n})$ *there exists an* n*-qubit quantum circuit with gate set* $\mathcal{G} = \{H, T, CNOT\}$ *that* ε*-approximately computes* U*.*

Recall that $T = S_{\pi/4}$, sometimes also referred to as $\pi/8$ gate. Up to an irrelevant global phase, it coincides with $R_Z(\pi/4)$.

Proof (idea). Doing the calculation, one can show that $THTH$ performs a rotation of the Bloch sphere with angle θ defined by $\cos(\theta/2) = \cos^2(\pi/8)$. This θ can be shown to be an irrational multiple of 2π; as a consequence, a rotation with *any* angle can be approximated by a suitable number of repetitions $THTH$. The same holds for $HTHT$, but with respect to a different axis. With these two rotations, it is then possible to do *any* rotation, and thus in particular the rotations R_Z and R_Y, and the claim then follows from Corollary 4.8. □

Theorem 4.10 (Solovay-Kitaev)**.** *Let* $\mathcal{G}$ *be a gate set that is closed under inversion, and let* $n \in \mathbb{N}$ *be a constant. Then, for any* $U \in \mathcal{U}(\mathcal{H}^{\otimes n})$ *that can be* ε*-approximately computed by a* n*-qubit quantum circuit with gate set* $\mathcal{G}$ *for any* $\varepsilon > 0$*, there exists an* n*-qubit quantum circuit with gate set* $\mathcal{G}$ *that* ε*-approximately computes* U *and has computational complexity* $O\big(\log^4(1/\varepsilon)\big)$*.*

In the remainder of these notes, when dealing with (efficient) quantum circuits, we typically leave the gate set $\mathcal{G}$ implicit, taking it as understood that the quantum circuits considered work with "simple" 1- and 2-qubit gates; this is well justified by the Solovay-Kitaev theorem.

4.5 Simon's Algorithm

As we have seen, the Deutsch-Jozsa algorithm performs exponentially better than any *deterministic* classical algorithm *in the worse case*, but only minorly better than a *randomized* classical algorithm with bounded-error. On the other hand, the Bernstein-Vazirani algorithm does clearly outperform any *randomized* classical algorithm, though only by a linear factor. Here, we present a problem where quantum algorithms are exponentially more efficient than randomized classical algorithms. We are still in the *query complexity* setting, where such a separation can be proven. Later, when discussing Shor's algorithm, we will see a super-polynomial separation in

computational complexity between quantum and classical algorithms, but those come without proofs due to the lack of classical lower-bound proofs.

Here, the computational problem is the following. Given a function $f : \{0,1\}^n \to \{0,1\}^n$ with the promise that there exists a non-zero "period" $s \in \{0,1\}^n$ such that

$$f(x) = f(x') \iff x' \in \{x, x \oplus s\} \tag{4.1}$$

for all $x, x' \in \{0,1\}^n$, find s. For a classical algorithm with black-box access to f, in order to find s, it must query f on two inputs x and x' with $x' = x \oplus s$ (or else must have excluded all other choices for s by means of such a pair of queries). Thus, an algorithm that has made q queries, and so can check $q(q-1)$ differences, has a probability $O(q^2/2^n)$ of having found s. For this to be, say, a constant, q needs to be $\Omega(2^{n/2})$, i.e., exponential in n. We emphasize that probabilities here are over the random choice over all functions with the given constraint. **Simon's algorithm** shows that one can do exponentially better with a quantum algorithm.

Proposition 4.11. *For any integers $n, k \in \mathbb{N}$, there exists a quantum algorithm with black-box access to $U_f \in \mathcal{U}(\mathcal{H}^{\otimes n} \otimes \mathcal{H}^{\otimes n})$ and query complexity $n+k-1$ that outputs s with property (4.1) with probability at least $1 - 2^{-k}$, assumed it exists.*

We remark that while the classical lower bound is meaningful only for a *randomly chosen* function f with the required property, the quantum upper bound holds for *any* such function.

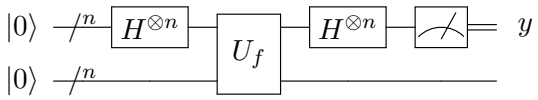


Figure 4.3: Quantum circuit for Simon's algorithm.

Proof. The algorithm is given in Figure 4.3. It starts off with $2n$ qubits in state $|0\rangle$ and applies H to the first n to obtain

$$\frac{1}{2^{n/2}} \sum_x |x\rangle|0\rangle \,,$$

where we understand that the sum is over $x \in \{0,1\}^n$. Applying U_f results in

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

For the purpose of this analysis, let us assume that we now measure the second half of the state; whether this measurement takes place now or later (or not at all) makes no difference in the distribution of k; this follows from the fact that actions on different registers commute.

As a result of this measurement, we observe some value $z \in \{0,1\}^n$, and the state of the first n qubits collapses to

$$\frac{1}{\sqrt{2}}(|x\rangle + |x+s\rangle)$$

where x is such that $f(x) = z$. Following previous patterns, the algorithm applies H again to the first n qubits. This results in

$$\frac{1}{2^{n/2}} \sum_y \frac{1}{\sqrt{2}}\big((-1)^{x\cdot y}|y\rangle + (-1)^{(x+s)\cdot y}|y\rangle\big) = \frac{1}{2^{n/2}} \sum_y \frac{1}{\sqrt{2}}(-1)^{x\cdot y}\big(1 + (-1)^{s\cdot y}\big)|y\rangle \,.$$

Now we observe that the amplitude of $|y\rangle$ is 0 if $s \cdot y = 1$, and it is $\pm 1/\sqrt{2^{n-1}}$ otherwise. Thus, by measuring, we will observe a uniformly random $y \in \{0,1\}^n$ with the property that $s \cdot y = 0$. By

repeating this procedure $n+k-1$ times, noting that $n+k-1$ random and independent vectors chosen from a vector space over $\mathbb{F}_2$ of dimension $n-1$ have full rank except with probability at most 2^{-k} (see below), s can then be found by means of basic linear algebra techniques. □

In the proof, we made use of the following technical lemma.

Lemma 4.12. *Let $\mathbb{F}$ be a finite field with cardinality q, and let V be a d-dimensional vector space over $\mathbb{F}$. For any integer $m \geq d$, consider the set*

$$\mathcal{NFR}_m(V) := \{(v_1, \ldots, v_m) \in V^m \mid \textit{the matrix } [v_1|\ldots|v_m] \textit{ has rank} < d\}\,.$$

Then, $pr_q(d,m) := \frac{|\mathcal{NFR}_m(V)|}{|V^m|} = \frac{|\mathcal{NFR}_m(V)|}{q^{dm}} \leq q^{d-m} - q^{-m}$.

We note that $pr_q(m,d)$ is well defined, given that $|\mathcal{NFR}_m(V)|$ only depends on the dimensison of V and the base field.

Proof. We prove the claim by induction. For the base case $d = 1$ and m arbitrary, we obviously have

$$pr_q(1,m) = q^{-m} \leq (q-1)q^{-m} = q^{1-m} - q^{-m}\,,$$

and for the base case $d = m$, we have that

$$pr_q(d,d) = 1 - (1-q^{-d})(1-q^{-d+1})\cdots(1-q^{-1}) \leq 1 - (1-q^{-1})^d \leq 1 - q^{-d}$$

using that $q \geq 2$, in the last step. For the induction step, we first note that if $v_m = 0$ then $(v_1, \ldots, v_m) \in \mathcal{NFR}_m(V)$ if and only if $(v_1, \ldots, v_{m-1}) \in \mathcal{NFR}_{m-1}(V)$. On the other hand, if $v_m \neq 0$ then it is easy to see that $(v_1, \ldots, v_m) \in \mathcal{NFR}_m(V)$ if and only if $(\bar{v}_1, \ldots, \bar{v}_{m-1}) \in \mathcal{NFR}_{m-1}(V/\langle v_m\rangle)$, where $\bar{v}_i$ is the redution of v_i mod $\langle v_m\rangle$. Therefore,

$$\begin{aligned} pr_q(d,m) &= q^{-d}\cdot pr_q(d,m-1) + (1-q^{-d})\cdot pr_q(d-1,m-1) \\ &\leq q^{-d}\cdot(q^{d-m+1} - q^{-m+1}) + (1-q^{-d})\cdot(q^{d-m} - q^{-m+1}) \\ &= q^{-m+1} + q^{d-m} - q^{-m} - q^{-m+1} \\ &= q^{d-m} - q^{-m}\,, \end{aligned}$$

which was to be proven. □

4.6 Grover's Algorithm for Unstructured Search

Grover's algorithm is again less impressive in terms of speed-up, but it applies to a very natural computational problem: given a function $f : \{0,1\}^n \to \{0,1\}$ for which there exist exactly M choices of $x \in \{0,1\}^n$ for which $f(x) = 1$, find such an x. Any classical algorithm with black-box access to f that succeeds with constant probability has query complexity $\Omega(2^n/M)$. With a quantum algorithm, we can gain a quadratic speed up.

Proposition 4.13. *For any positive integers n and $M \leq 2^n$, there exists a quantum algorithm with black-box access to $U_f \in \mathcal{U}(\mathcal{H}^{\otimes n} \otimes \mathcal{H})$ and query complexity $O\big(\sqrt{2^n/M}\big)$ that outputs x with $f(x) = 1$ with probability at least $1 - M/2^n$, given that there exist exactly M such x's.*

Note that the algorithm needs to know M, which is not very realistic in typical applications. As will become from the proof, the algorithm still works though, say with success probability at least $1/2$, if a sufficiently good approximation is known. As a matter of fact, by essentially running the algorithm with cleverly chosen guesses for M, an expected number of $O\big(\sqrt{2^n/M}\big)$ queries still suffice to find a solution in case M is unknown.

Proof. By the phase kickback technique from Sections 4.1 and 4.2, we may just as well assume black box access to $V_f \in \mathcal{U}(\mathcal{H}^{\otimes n})$ with $V_f|x\rangle = (-1)^{f(x)}|x\rangle$. The algorithm starts off with $|0\rangle \in \mathcal{S}(\mathcal{H}^{\otimes n})$ and applies $H^{\otimes n}$. This is followed by ℓ *Grover iterations* and, finally, the resulting state is measured. As illustrated in Figure 4.4 below, the Grover iteration consists of: applying V_f, applying $H^{\otimes n}$, applying a conditional phase shift $P = 2|0\rangle\langle 0| - \mathbb{I} \in \mathcal{U}(\mathcal{H}^{\otimes n})$, which is such that $P|0\rangle = |0\rangle$ and $P|x\rangle = -|x\rangle$ for $x \neq 0$, and applying $H^{\otimes n}$ once more.

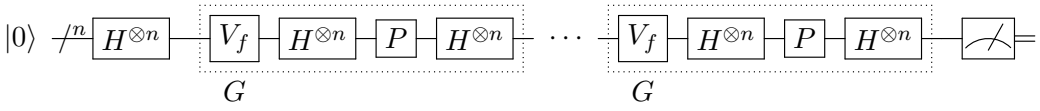


Figure 4.4: Quantum circuit for Grover's algorithm.

We now analyze the algorithm and determine ℓ, which then obviously determines the query complexity. Consider

$$|\psi\rangle = H^{\otimes n}|0\rangle = \frac{1}{2^{n/2}} \sum_x |x\rangle \,,$$

and note that $H^{\otimes n} P H^{\otimes n} = H^{\otimes n}(2|0\rangle\langle 0| - \mathbb{I})H^{\otimes n} = 2|\psi\rangle\langle\psi| - \mathbb{I}$, so that the Grover may be written as

$$G = H^{\otimes n} P H^{\otimes n} V_f = (2|\psi\rangle\langle\psi| - \mathbb{I})V_f \,.$$

On order to understand the action of G, we set

$$|good\rangle = \frac{1}{\sqrt{M}} \sum_{\substack{x \text{ s.t.} \\ f(x)=1}} |x\rangle \qquad \text{and} \qquad |bad\rangle = \frac{1}{\sqrt{N-M}} \sum_{\substack{x \text{ s.t.} \\ f(x)=0}} |x\rangle \,,$$

where $N = 2^n$, and we observe that we can write

$$|\psi\rangle = \sqrt{\frac{N-M}{N}}\,|bad\rangle + \sqrt{\frac{M}{N}}\,|good\rangle = \cos(\theta_\circ)\,|bad\rangle + \sin(\theta_\circ)\,|good\rangle =: |\psi_{\theta_\circ}\rangle$$

for the proper choice of $\theta_\circ \in [0, \frac{\pi}{2}]$. Furthermore, for an arbitrary θ, we have that

$$V_f|\psi_\theta\rangle = \cos(\theta)\,|bad\rangle - \sin(\theta)\,|good\rangle = |\psi_{-\theta}\rangle \,.$$

Furthermore, by basic trigonometric identities,

$$\begin{aligned}\big(2|\psi\rangle\langle\psi| - \mathbb{I}\big)|bad\rangle &= (2\cos(\theta_\circ)^2 - 1)\,|bad\rangle + 2\sin(\theta_\circ)\cos(\theta_\circ)\,|good\rangle \\ &= \cos(2\theta_\circ)\,|bad\rangle + \sin(2\theta_\circ)\,|good\rangle\end{aligned}$$

and

$$\begin{aligned}\big(2|\psi\rangle\langle\psi| - \mathbb{I}\big)|good\rangle &= 2\sin(\theta_\circ)\cos(\theta_\circ)\,|bad\rangle + (2\sin(\theta_\circ)^2 - 1)\,|good\rangle \\ &= \sin(2\theta_\circ)\,|bad\rangle - \cos(2\theta_\circ)\,|good\rangle \,.\end{aligned}$$

Putting all together, we obtain

$$\begin{aligned}G|\psi_\theta\rangle &= \big(2|\psi\rangle\langle\psi| - \mathbb{I}\big)V_f|\psi_\theta\rangle \\ &= \big(2|\psi\rangle\langle\psi| - \mathbb{I}\big)\big(\cos(\theta)\,|bad\rangle - \sin(\theta)\,|good\rangle\big) \\ &= \big(\cos(\theta)\cos(2\theta_\circ) - \sin(\theta)\sin(2\theta_\circ)\big)\,|bad\rangle + \big(\cos(\theta)\sin(2\theta_\circ) + \sin(\theta)\cos(2\theta_\circ)\big)\,|good\rangle \\ &= \cos(\theta + 2\theta_\circ)\,|bad\rangle + \sin(\theta + 2\theta_\circ)\,|good\rangle \,.\end{aligned}$$

As such, one application of G simply increases the angle by $2\theta_\circ$. Therefore, the state $|\psi\rangle = |\psi_{\theta_\circ}\rangle$ after the application of the inital $H^{\otimes n}$ evolves as $|\psi_{\theta_\circ}\rangle, |\psi_{3\theta_\circ}\rangle, |\psi_{5\theta_\circ}\rangle$ etc. Ideally, we want to choose ℓ such that $(2\ell+1)\theta_\circ = \frac{\pi}{2}$, so that the final state that is measured equals $|good\rangle$, and so we observe x with $f(x) = 1$ with certainty. Furthermore, using that $\sqrt{M/N} = \sin(\theta_\circ) \leq \theta_\circ$, we then have that ℓ, and thus the query complexity of the algorithm, is $O(1/\theta_\circ) = O(\sqrt{N/M})$. In general, this choice of ℓ will not be an integer, and then we have to round to the closest integer; the probability of observing x with $f(x) = 1$ is then still at least

$$\sin\Big(\frac{\pi}{2} \pm \theta_\circ\Big)^2 = 1 - \cos\Big(\frac{\pi}{2} \pm \theta_\circ\Big)^2 = 1 - \sin(\theta_\circ)^2 = 1 - M/N\,,$$

as claimed. □

Our focus here is on query complexity, but we do want to point out that by means of the techniques from Section 2.5, the P gate can be implemented using $O(n)$ elementary gates (plus $O(n)$ work qubits), and thus the computational complexity is larger by a factor $O(n)$ only.

Chapter 5

(More) Quantum Algorithms Based on the Fourier Transform

The goal of this chapter will be to explain and analyze Shor's quantum algorithms for factoring large integers and for computing discrete logarithms. These algorithms run in polynomial time (in the description length of the problem), in contrast to the best-known classical algorithms (like the number-field sieve), which run in superpolynomial time.[1] These quantum algorithms are particularly exciting — or troubling, depending on the point of view — given that the security of (almost) all of currently used public-key cryptography relies on the assumed hardness of these two computational problems. This means that a scalable quantum computer that is capable of running Shor's algorithms would render today's internet completely insecure.

We start the chapter by introducing the Fourier transform, which is at the core of Shor's algorithms, and we then first discuss some algorithms that avoid some of the technical difficulties but still reflect the basic idea behind Shor's algorithms. As a matter of fact, at the heart of the chapter will be a meta quantum algorithm for solving the so-called *hidden subgroup problem* for Abelian groups; almost all quantum algorithms we treat in these notes can be understood as an instantiation of the meta algorithm up to some modifications.

5.1 The Classical and the Quantum Fourier Transform

Fix an integer $N \geq 2$ and consider the ring $\mathbb{Z}/N\mathbb{Z}$. By default, the elements are represented by integers in $\{0, \ldots, N-1\}$, and so we will typically not distinguish between an integer and its coset modulo N. We let $\omega_N := e^{2\pi i/N} \in \mathcal{S}(\mathbb{C}) \subset \mathbb{C}$ and write ω when N is clear.[2] Recall that, by Euler's identity, the function $\mathbb{Z} \to \mathbb{C}$, $j \mapsto \omega_N^j = e^{2\pi ij/N}$ is a group homomorphism with kernel $N\mathbb{Z}$, and so we may also understand it as a function $\mathbb{Z}/N\mathbb{Z} \to \mathbb{C}$. By allowing j to be real valued, we can also consider it as a function $\mathbb{R} \to \mathbb{C}$, or $\mathbb{R}/N\mathbb{Z} \to \mathbb{C}$.

Definition 5.1. *For any function $\alpha : \mathbb{Z}/N\mathbb{Z} \to \mathbb{C}$ the (discrete)* **Fourier transform** *of α is the function $\hat{\alpha} : \mathbb{Z}/N\mathbb{Z} \to \mathbb{C}$ given by*

$$\hat{\alpha}(k) = \frac{1}{\sqrt{N}} \sum_{j=0}^{N-1} \omega_N^{jk} \alpha(j)$$

[1] In case of computing discrete logarithms, the running time of classical algorithms depends on the group considered and ranges from polynomial time (for "easy" groups) to exponential time (e.g. for elliptic curves).

[2] As is common, we use e^z as a shorthand for $\exp(z)$. This is consistent with the basic property of the exponential function that $\exp(z+w) = \exp(z)\exp(w)$, and, thus, $\exp(n \cdot z) = \exp(z)^n$ for any integer n.

for any $k \in \mathbb{Z}/N\mathbb{Z}$.[3]

We may also write α_j instead of $\alpha(j)$ and $\hat{\alpha}_k$ instead of $\hat{\alpha}(k)$.

We quickly look at a simple yet important example. The Fourier transform of the all-1 function $1_N : \mathbb{Z}/N\mathbb{Z} \to \mathbb{C}$, $j \mapsto 1$ is given by

$$\hat{1}_N(k) = \frac{1}{\sqrt{N}} \sum_{j=0}^{N-1} \omega_N^{jk} = \begin{cases} \sqrt{N} & \text{if } k = 0 \\ 0 & \text{else} \end{cases}$$

for $k \in \mathbb{Z}/N\mathbb{Z}$, sometimes referred to as **Dirac function**.

We point out, and this will become important later on, that the definition of $\hat{\alpha} : \mathbb{Z}/N\mathbb{Z} \to \mathbb{C}$ above extends to $\hat{\alpha} : \mathbb{R}/N\mathbb{Z} \to \mathbb{C}$, simply by allowing k in Definition 5.1 to be in the ring $\mathbb{R}/N\mathbb{Z}$. In particular, we will consider $\hat{1}_N(k)$ as above but for real-valued k. The function $\hat{1}_N(k)$ will then not necessarily vanish for non-zero k (see Figure 5.1); what will be important for us that that in the neighbourhood of 0, it is still relatively large, as captured by the following.

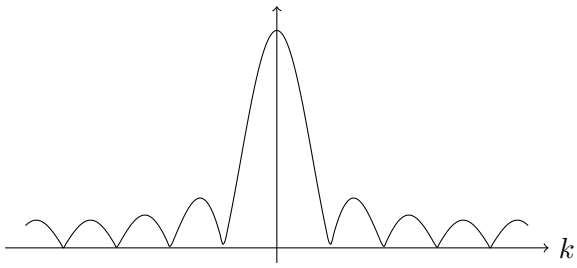


Figure 5.1: The absolute value of $\hat{1}_N$ as a function $\mathbb{R}/N\mathbb{Z} \to \mathbb{C}$.

Lemma 5.1. *As function* $\hat{1}_N : \mathbb{R}/N\mathbb{Z} \to \mathbb{C}$*, and for* $\xi \in \mathbb{R}$ *with* $|\xi| \leq 1/2$*,*

$$\left|\hat{1}_N(\xi)\right| \geq \frac{2\sqrt{N}}{\pi} = \frac{2}{\pi} \cdot \hat{1}_N(0)\,.$$

Proof. We observe that

$$\hat{1}_N(\xi) = \frac{1}{\sqrt{N}} \sum_{j=0}^{N-1} \omega^{j\xi} = \frac{1}{\sqrt{N}} \sum_{j=0}^{N-1} \left(\omega^{\xi}\right)^j = \frac{1}{\sqrt{N}} \frac{1-\omega^{\xi N}}{1-\omega^{\xi}} = \frac{1}{\sqrt{N}} \frac{1-e^{2\pi i \xi}}{1-e^{2\pi i \xi/N}}\,.$$

where the last equality is the well-known closed form of a geometric series, assuming that $\xi \neq 0$. From the goemetric picture, by comparing chord and arc, it is easy to convince yourself that $2\pi\nu \geq |1 - e^{2\pi i \nu}| \geq 4\nu$ for $\nu \in [-1/2, 1/2]$, so that

$$\left|\hat{1}_N(\xi)\right| \geq \frac{1}{\sqrt{N}} \frac{4\xi}{2\pi\xi/N} = \frac{2\sqrt{N}}{\pi}\,,$$

as claimed. □

We now introduce the *quantum* Fourier transform. Given the integer $N \geq 2$, we consider an arbitrary N-dimensional Hilbert space $\mathcal{H}_N$ with an orthonormal basis $\{|0\rangle, \ldots, |N-1\rangle\}$, where the basis vectors are labelled by the integers in $\{0, \ldots, N-1\}$, which we identify with the elements in $\mathbb{Z}/N\mathbb{Z}$. The quantum Fourier transform is simply the classical Fourier transform applied to the amplitudes.

[3]The deviation from the typical definition, which uses the conjugate coefficients $e^{-2\pi i jk/N}$, is an artefact of having inner products conjugate-linear in the *first* argument, rather than in the second.

Definition 5.2. *The* **quantum Fourier transform** *on $\mathcal{H}_N$ (w.r.t. $\{|j\rangle\}_{j\in\{0,\dots,N-1\}}$) is the unitary operator $F \in \mathcal{U}(\mathcal{H}_N)$ given by*

$$F : \sum_{j=0}^{N-1} \alpha_j |j\rangle \mapsto \sum_{k=0}^{N-1} \hat{\alpha}_k |k\rangle \,.$$

In other words,

$$F : |j\rangle \mapsto \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} \omega_N^{jk} |k\rangle$$

for any $j \in \{0, \dots, N-1\}$.

As for F being unitary, we observe that for any pair $|m\rangle$ and $|k\rangle$ of basis vectors,

$$\langle m|F^\dagger F|k\rangle = \frac{1}{N} \sum_{j,\ell} \omega^{jk-\ell m} \langle \ell | j \rangle = \frac{1}{N} \sum_{j} \omega^{j(k-m)} = \frac{1}{\sqrt{N}} \hat{1}_N(k-m) = \langle m|k\rangle \,.$$

We point out that mathematically, the quantum Fourier transform is identical to the ordinary Fourier transform, except that it is made explicit that the functions $\alpha : \mathbb{Z}/N\mathbb{Z} \to \mathbb{C}$ form a Hilbert space, and Dirac's bra-ket notation is used. Furthermore, when it comes to *computing* the quantum Fourier transform, a different model of computation is used.

5.2 Period Finding in $\mathbb{Z}/N\mathbb{Z}$

As a first application, we show here that the quantum Fourier transform gives rise to a variant of Simon's algorithm. Consider a black-box function $f : \mathbb{Z}/N\mathbb{Z} \to \mathcal{R}$ with an arbitrary finite domain $\mathcal{R}$, with the promise that there exists a non-zero divisor s of N such that $f(x) = f(x')$ if and only if s divides $x - x'$. The goal is to find s. Looking ahead, Shor's factoring algorithm will also work by finding the period of some function; however, there and in contrast to here, no multiple of the period will be given.

We show that, very much in spirit of Simon's algorithm, the following quantum circuit will allow us to obtain useful information on s. Note that the state $|0\rangle$ of the first register is given by the first basis vector of the fixed basis for $\mathcal{H}_N$. On the other hand, the state of the second register, which is also denoted $|0\rangle$, is a quantum encoding of the neutral element of $\mathcal{R}$ when considered as a group, as done for defining U_f.

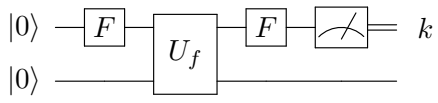


Figure 5.2: Quantum circuit for period-finding on $\mathbb{Z}/N\mathbb{Z}$.

Proposition 5.2. *Given that f is as promised, the quantum circuit in Figure 5.2 produces a uniformly random $k \in \mathbb{Z}/N\mathbb{Z}$ subject to $k \cdot s = 0$ (in $\mathbb{Z}/N\mathbb{Z}$).*

Note that $ks = 0$ in $\mathbb{Z}/N\mathbb{Z}$ means that $ks = cN$ over the integers, for a random integer $c \in \{0, \dots, s-1\}$. In order to obtain s, we can then bring the fraction k/N, which equals c/s, into reduced form and read out the denominator; this results in $s' = s/\gcd(c, s)$. Thus, if c happens to be coprime, we have $s' = s$, and we can easily check whether $s' = s$. Furthermore, by running the procedure twice, we obtain s by taking the least common multiple of the two

corresponding s''s *if* the two corresponding c's are coprime. The latter can be shown to happen with probability at least[4] $2-\pi^2/6 \approx 0.35$ and approaches $1/\zeta(2) = 6/\pi^2 \approx 0.6$ for large s. Thus, the correct s can be obtained with overhelming probability by repeating the above procedure a few times.

Proof. Applying F to the first register, and then appling U_f, results in

$$|0\rangle|0\rangle \mapsto \frac{1}{\sqrt{N}} \sum_{x=0}^{N-1} |x\rangle|0\rangle \mapsto \frac{1}{\sqrt{N}} \sum_{x=0}^{N-1} |x\rangle|f(x)\rangle \,.$$

For the purpose of the analysis, let as assume that the second register is now measured. The measurement outcome is a element $y \in \mathcal{R}$, and the state of the first register collapses to a superposition of all $x \in \mathbb{Z}/N\mathbb{Z}$ with $f(x) = y$. By assumption on f, this means that the state of the first register becomes

$$\frac{1}{\sqrt{m}} \sum_{\ell=0}^{m-1} |\ell s + x\rangle \,,$$

where x is some $x \in \mathbb{Z}/N\mathbb{Z}$ with $f(x) = y$, and $m := N/s$. Applying F then results in

$$\begin{aligned}
\frac{1}{\sqrt{Nm}} \sum_{\ell=0}^{m-1} \sum_{k=0}^{N-1} \omega_N^{(\ell s+x)k} |k\rangle &= \frac{1}{\sqrt{Nm}} \sum_{k=0}^{N-1} \sum_{\ell=0}^{m-1} \omega_N^{\ell sk} \omega_N^{xk} |k\rangle \\
&= \frac{1}{\sqrt{Nm}} \sum_{k=0}^{N-1} \Bigg(\sum_{\ell=0}^{m-1} \omega_m^{\ell k} \Bigg) \omega_N^{xk} |k\rangle \\
&= \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} \hat{1}_m(k)\, \omega_N^{xk} |k\rangle \,.
\end{aligned}$$

where we used that $\omega_N^s = e^{2\pi i s/N} = e^{2\pi i/m} = \omega_m$. From this, given that $\hat{1}_m(k)$ does not vanish if and only if k is an integer multiple of $m = N/s$, we see that measuring this register results in a uniformly random $k \in \mathbb{Z}/N\mathbb{Z}$ subject to $k \cdot s = 0$. □

5.3 Period Finding in $\mathbb{Z}/N\mathbb{Z} \times \mathbb{Z}/N\mathbb{Z}$

We consider a small variation of the period finding problem from Section 5.2 above, where now $f : \mathbb{Z}/N\mathbb{Z} \times \mathbb{Z}/N\mathbb{Z} \to \mathcal{R}$, and it is promised that there exists a non-zero $s \in \mathbb{Z}/N\mathbb{Z}$ such that $f(x, y) = f(x', y')$ if and only if $x - x' = s(y - y')$, i.e., $(s, 1)$ divides $(x, y) - (x', y')$. We remark that restricting the second component of the period to be 1, or, equivalently, to be co-prime with N, is merely for simplicity. Looking ahead, we point out that applied to the function $(x, y) \mapsto g^x h^y$ for two group elements g, h, finding s is equivalent to computing the discrete logarithm of h with respect to g.

Proposition 5.3. *Assuming that $f : \mathbb{Z}/N\mathbb{Z} \times \mathbb{Z}/N\mathbb{Z} \to \mathcal{R}$ is such that there exists a non-zero $s \in \mathbb{Z}/N\mathbb{Z}$ such that $f(x, y) = f(x', y')$ if and only if $x - x' = s(y - y')$, the quantum circuit in Figure 5.3 produces uniformly random $k, m \in \mathbb{Z}/N\mathbb{Z}$ subject to $sk + m = 0$ (in $\mathbb{Z}/N\mathbb{Z}$).*

Similarly to before, if we are unlucky and k and N are not coprime then the procedure needs to be repeated, and with high probability s will be recovered after a few repetitions.

[4]Indeed, the probability of the two c's to share a common factor is at most $\sum_{n \geq 2} 1/n^2 = \zeta(2) - 1 = \pi^2/6 - 1$.

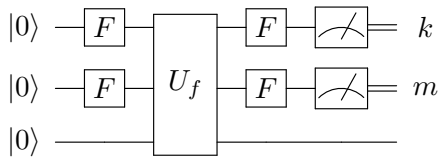


Figure 5.3: Quantum circuit for period-finding on $\mathbb{Z}/N\mathbb{Z}$.

Proof. Applying F to the first two registers, and then appling U_f, results in

$$|0\rangle|0\rangle|0\rangle \mapsto \frac{1}{N}\sum_{x,y=0}^{N-1}|x\rangle|y\rangle|0\rangle \mapsto \frac{1}{N}\sum_{x,y=0}^{N-1}|x\rangle|y\rangle|f(x,y)\rangle\,.$$

Again, let us assume that the last register is now measured. The measurement outcome is a element $z \in \mathcal{R}$, and the state of the first two registers collapses to a superposition of all x and y in $\mathbb{Z}/N\mathbb{Z}$ with $f(x,y) = z$. By the assumption on f, this means that the state of the first two registers becomes

$$\frac{1}{\sqrt{N}}\sum_{\ell=0}^{N-1}|\ell s + x\rangle|\ell + y\rangle\,,$$

where x and y are so that with $f(x,y) = z$. Applying F individually to both registers then results in

$$\begin{aligned}\frac{1}{N^{3/2}}\sum_{\ell,k,m}\omega^{(\ell s+x)k}\omega^{(\ell+y)m}|k\rangle|m\rangle &= \frac{1}{N^{3/2}}\sum_{\ell,k,m}\omega^{\ell(sk+m)}\omega^{xk}\omega^{ym}|k\rangle|m\rangle \\ &= \frac{1}{N}\sum_{k,m}\hat{1}_N(sk+m)\,\omega^{xk}\omega^{ym}|k\rangle|m\rangle\,.\end{aligned}$$

From this we see that measuring these two registers results in uniformly random $k, m \in \mathbb{Z}/N\mathbb{Z}$ subject to $sk + m = 0$, as claimed. □

5.4 The Hidden Subgroup Problem

The reader should start to see a pattern. Indeed, the computational problems solved by Deutsch's algorithm, by Bernstein-Vazirani's algorithm, and by Simon's algorithm, as well as the two period finding problems above, are all instances of the **hidden subgroup problem (HSP)**. In its most generality, the HSP reads as follows. Given a group G and a black-box function $f : G \to \mathcal{R}$ with the promise that $f(x) = f(x')$ if and only if $x^{-1}x' \in H$, where H is an unknown subgroup of G, the goal is to find (a representation of) the "hidden subgroup" H. In other words, f acts identically within each coset xH, but differently on different cosets.

For Deutsch's problem, we have $G = \mathbb{Z}/2\mathbb{Z}$ and H is either $\mathbb{Z}/2\mathbb{Z}$ or $\{0\}$. For Bernstein-Vazirani's problem and for Simon's problem, we have $G = \mathbb{F}_2^n = \mathbb{F}_2 \times \cdots \times \mathbb{F}_2$, understood as vector space over the binary field $\mathbb{F}_2$ (with standard inner product), and H is the orthogonal complement of $a \in G$ for the former, and the linear span of (non-zero) $s \in G$. For period finding in $\mathbb{Z}/N\mathbb{Z}$, we have $G = \mathbb{Z}/N\mathbb{Z}$ and H is generated by (non-zero) $s \in G$. And, eventually, for period finding in $\mathbb{Z}/N\mathbb{Z} \times \mathbb{Z}/N\mathbb{Z}$, we have $G = \mathbb{Z}/N\mathbb{Z} \oplus \mathbb{Z}/N\mathbb{Z}$ and H generated by $(s, 1) \in G$.

Not so suprising, the techniques for solving the above specific problems by means of quantum algorithms unify and generalize. We note, however, that Deutsch's algorithm and Bernstein-Vazirani's algorithm are both slighty different than the generic hidden subgroup algorithm

outlined below; indeed, they produce the right answer in one go and with certainty. Similarly, the Deutsch-Jozsa problem does not fit into the HSP framework (unless $n = 1$).

We start with the following definition.

Definition 5.3. *For a finite Abelian group G, the* **dual group** $\hat{G}$ *is the group of homomorphisms χ from G to the multiplicative group $\mathcal{S}(\mathbb{C})$. $\chi \in \hat{G}$ is called a* **character**.

The following captures some basic properties of the dual group.

Proposition 5.4. $|\hat{G}| = |G|$. *Furthermore, the* **orthogonality relations**

$$\sum_x \overline{\chi(x)}\chi'(x) = \begin{cases} 0 & \text{if } \chi \neq \chi' \\ |G| & \text{if } \chi = \chi' \end{cases} \quad \text{and} \quad \sum_\chi \overline{\chi(x)}\chi(x') = \begin{cases} 0 & \text{if } x \neq x' \\ |\hat{G}| & \text{if } x = x' \end{cases}$$

hold for all $x, x' \in G$ and $\chi, \chi' \in \hat{G}$. In particular, for every $1 \neq x \in G$ and $1 \neq \chi \in \hat{G}$,

$$\sum_{x \in G} \chi(x) = 0 \qquad \text{and} \qquad \sum_{\chi \in \hat{G}} \chi(x) = 0\,.$$

Proof. We start by observing that $\hat{G}$ must be finite, given that G is. This follows from the fact that $\chi(g)^{|G|} = \chi(g^{|G|}) = \chi(1) = 1$, and thus $\chi(g)$ must be a $|G|$-th root of unity for any g.

As for the last two claims of the statement, the first one follows from the observation that

$$\sum_x \chi(x) = \sum_x \chi(xy) = \chi(y) \sum_x \chi(x)$$

for any $y \in G$, and the existence of $y \in G$ with $\chi(y) \neq 1$. Similarly for the other one, using that $|\hat{G}| < \infty$, and that for any $x \neq 1$ there exists $\chi \in \hat{G}$ with $\chi(x) \neq 1$ (which follows from Lemma 5.5 below). Since χ maps into $\mathcal{S}(\mathbb{C})$, we have that $\chi^{-1} = \overline{\chi}$, and the two orthogonality relations then follow easily. Also, the orthogonality relations then imply that $|\hat{G}| = |G|$. □

Lemma 5.5. *Let $H \subsetneq G$ be a subgroup of G and $x \in G \setminus H$, and let $\pi \in \hat{H}$ be a character of H. Then there exists a character $\chi \in \hat{G}$ so that $\chi(h) = \pi(h)$ for all $h \in H$, and $\chi(x) \neq 1$.*

Proof. By repeated application, it is enough to show the claim in case G is generated by H and x, i.e., $G = \{x^i h \,|\, h \in H,\, i \in \mathbb{Z}\}$. For that, let $n \geq 2$ is the smallest positive integer for which $x^n \in H$, and let $\omega \neq 1$ be such that $\omega^n = \pi(x^n)$. For any $g = x^i h \in G$, we now define $\chi(g) = \chi(x^i h) := \omega^i \pi(h)$. We first need to show that this is well defined. Let $h, h' \in H$ and $i, j \in \mathbb{Z}$ with $x^i h = x^j h'$. Then $x^{i-j} = h/h' \in H$, and thus $i - j = kn$ for $k \in \mathbb{Z}$. Therefore,

$$\chi(x^i h)/\chi(x^j h') = \omega^{i-j}\chi(h/h') = \omega^{kn}\pi(h/h') = \pi(x^n)^k \pi(h/h') = \pi(x^{kn}h/h') = \pi(1) = 1\,.$$

Finally, it is obvious that χ is a homomorphism, i.e., $\chi \in \hat{G}$, and $\chi(x) = \omega \neq 1$. □

As an immediate consequence, for any pair of fixed orthonormal bases $\{|x\rangle\}_{x \in G}$ and $\{|\chi\rangle\}_{\chi \in \hat{G}}$ of the Hilbert space $\mathbb{C}^{|G|}$, the **generalized quantum Fourier transform** F, given as follows, is unitary:

$$F : |x\rangle \mapsto \frac{1}{\sqrt{|G|}} \sum_{\chi \in \hat{G}} \chi(x)|\chi\rangle\,.$$

In case of $G = \mathbb{Z}/N\mathbb{Z}$, the characters are the homomorphisms that map j to ω_N^{jk} for k in $\{0, \ldots, N-1\}$, and so F coincides with Definition 5.2. In case of $G = \mathbb{Z}/2\mathbb{Z} \oplus \cdots \oplus \mathbb{Z}/2\mathbb{Z}$, the characters are the functions $x \mapsto (-1)^{x \cdot y}$ for all possible choices of y, and then F matches $H^{\otimes n}$. This allows us to generalize above quantum algorithms and obtain the following general result.

Theorem 5.6. *For any finite Abelian group G with hidden subgroup H, there exists a quantum circuit with black-box access to U_f and query complexity 1 that outputs a random character χ that acts trivially on H.*

In general, in order to uniquely identify H, the algorithm needs to be repeated several times; how (a representation of) H can then be computed depends on the structure of G and H.

Proof. We produce a superposition of all $x \in G$ (e.g., by applying $F^\dagger$ to $|1\rangle$) and an ancilla and apply U_f to get

$$\frac{1}{\sqrt{|G|}} \sum_{x \in G} |x\rangle|0\rangle \mapsto \frac{1}{\sqrt{|G|}} \sum_{x \in G} |x\rangle|f(x)\rangle \,.$$

Measuring the second register to observe $y = f(x)$ has the effect that the state collapses to

$$\frac{1}{\sqrt{|H|}} \sum_{h \in H} |xh\rangle \,.$$

Applying the generalized quantum Fourier transform F then results in

$$\frac{1}{\sqrt{|G||H|}} \sum_{h \in H} \sum_{\chi \in \hat{G}} \chi(xh)|\chi\rangle = \frac{1}{\sqrt{|G||H|}} \sum_{\chi \in \hat{G}} \chi(x) \sum_{h \in H} \chi(h)\,|\chi\rangle$$

and so, given that the restriction of χ to H is a character in $\hat{H}$, and thus Proposition 5.4 applies correspondingly, we see that $|\chi\rangle$ has a non-trivial amplitude only if χ acts trivially on H. □

Later, we show that the technique even generalizes to *infinite* Abelian groups to some extent. This will then eventually lead to Shor's famous factoring algorithm.

5.5 Computing the Quantum Fourier Transform

We have seen that the quantum Fourier transform is very powerful and gives rise to various quantum algorithms with low *query* complexity. In order to obtain quantum algorithms with low *computational* complexity, we need to be able to efficiently compute the quantum Fourier transform. We show here how this is done.

Note that in order to fit into our model of computation, we need F to act on qubits. As such, we set $N = 2^n$ for a positive integer n, and we let $\mathcal{H}_N$ be $\mathcal{H}_N = \mathcal{H}^{\otimes n} = \mathcal{H} \otimes \cdots \otimes \mathcal{H}$. Finally, we require the basis $\{|0\rangle, \ldots, |N-1\rangle\}$ to be given by $|j\rangle = |j_1\rangle \cdots |j_n\rangle$, where $(j_1, \ldots, j_n) \in \{0,1\}^n$ is the **binary representation** of $j \in \{0, \ldots, N-1\}$, uniquely determined by the equality

$$j = [j_1 \cdots j_n] := \sum_{\ell=1}^{n} j_\ell \, 2^{n-\ell} \,.$$

By a variation of this formalism, we also write

$$[0.j_1 \cdots j_n] := j/2^n = \sum_{\ell=1}^{n} j_\ell \, 2^{-\ell}$$

The following observation is at the core of the efficient computability of the quantum Fourier transform F, which we will denote by F_n below, in order to make the dependency on the parameter n explicit.

Lemma 5.7. *For any positive integer n and any $j = [j_1 \cdots j_n] \in \{0, \ldots, 2^n - 1\}$:*

$$F_n|j\rangle = F_n|j_1\rangle \cdots |j_n\rangle = \frac{1}{\sqrt{2^n}} \bigotimes_{\ell=1}^{n} \Big(|0\rangle + \omega_{2^\ell}^{j}|1\rangle\Big),$$

or, in other words,

$$= \frac{1}{\sqrt{2^n}} \Big(|0\rangle + e^{2\pi i[0.j_n]}|1\rangle\Big) \Big(|0\rangle + e^{2\pi i[0.j_{n-1}j_n]}|1\rangle\Big) \cdots \Big(|0\rangle + e^{2\pi i[0.j_1 \cdots j_{n-1}j_n]}|1\rangle\Big).$$

Proof. By basic term manipulations, we see that

$$\begin{aligned}
F_n|j\rangle &= \frac{1}{\sqrt{2^n}} \sum_k e^{2\pi i jk/2^n} |k_1\rangle \cdots |k_n\rangle \\
&= \frac{1}{\sqrt{2^n}} \sum_{k_1 \ldots k_n} \prod_\ell e^{\pi i j k_\ell / 2^{n-\ell}} |k_1\rangle \cdots |k_n\rangle \\
&= \frac{1}{\sqrt{2^n}} \sum_{k_1 \ldots k_n} \bigotimes_\ell e^{2\pi i j k_\ell / 2^\ell} |k_\ell\rangle \\
&= \frac{1}{\sqrt{2^n}} \bigotimes_\ell \sum_{k_\ell} e^{2\pi i j k_\ell / 2^\ell} |k_\ell\rangle \\
&= \frac{1}{\sqrt{2^n}} \bigotimes_\ell \Big(|0\rangle + e^{2\pi i j / 2^\ell}|1\rangle\Big).
\end{aligned}$$

This shows the claim. The second variant follows from the fact that $x \mapsto e^{2\pi i x}$ is cyclic with period 1, and noting that $j/2^\ell \bmod 1 = [0.j_{n-\ell+1} \cdots j_n]$. □

Recalling the definition of the phase shift gate, the above shows that the first qubit of $F_n|j_1\rangle \cdots |j_n\rangle$ equals $S_{2\pi[0.j_n]}|+\rangle$, and the remaining qubits are given by

$$S_{2\pi[0.0j_n]}\Big(|0\rangle + e^{2\pi i[0.j_{n-1}]}|1\rangle\Big) \otimes \cdots \otimes S_{2\pi[0.0\cdots 0j_n]}\Big(|0\rangle + e^{2\pi i[0.j_1 \cdots j_{n-1}]}|1\rangle\Big).$$

This proves the following.

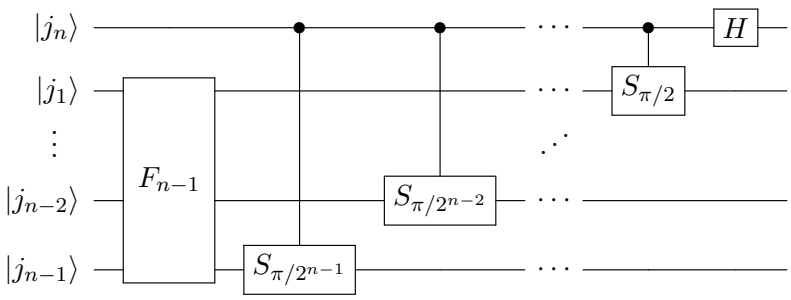
Corollary 5.8. *For any integer $n > 1$ and any $j_1, \ldots, j_n \in \{0, 1\}$:*

$$F_n|j_1\rangle \cdots |j_n\rangle = \big(S_{\pi j_n} \otimes S_{\pi j_n/2} \otimes \cdots \otimes S_{\pi j_n/2^{n-1}}\big)\big(|+\rangle \otimes F_{n-1}|j_1\rangle \cdots |j_{n-1}\rangle\big).$$

Thus, F_n can be recursively computed. In particular, the following holds.

Theorem 5.9. *The n-qubit quantum Fourier transform F_n can be computed with computational complexity $O(n^2)$ with a quantum circuit with gate set consisting of H, swap gates, and controlled phase shift gates. The same holds for the inverse $F_n^\dagger$.*

Proof. Given the above corollary, it follows by trivial inspection that, up to the order of the wires, F_n can be recursively computed by means of the following quantum circuit.



This proves the claim on F_n. For the inverse, we observe that we can invert the circuit for F_n by running the circuit "backwards" and that the considered gate set is preserved under inversion. □

5.6 Shor's Algorithm for Period Finding in $\mathbb{Z}$

Here, we consider a black-box function $f : \mathbb{Z} \to \mathcal{R}$ with an arbitrary finite domain $\mathcal{R}$, with the promise that there exists $s \in \mathbb{Z}$ in the range $1 < s \leq N = 2^n$ such that $f(x) = f(x')$ if and only if $x - x'$ is an integer multiple of s. The goal is to find s.

The crucial difference to Section 5.2 is that, here, s does not divide N, and therefore f does not naturally induce a function $\mathbb{Z}/N\mathbb{Z} \to \mathcal{R}$. Therefore, if we apply the Fourier transform to $f : \{0, \ldots, N-1\} \to \mathcal{R}$, the wrap-around we get by taking numbers modulo N messes up things. Fortunately, we will be able to control how much things are messed up. As a starting point, let us simply take the circuit from Figure 5.2 in order to try to solve the above variation of the period-finding problem, and we follow the analysis from Section 5.2 as far as we can.

Applying F to the first register, which is now an n-qubit register, applying U_f, and then measuring the second register produces a state of the form

$$\frac{1}{\sqrt{m}} \sum_{\ell=0}^{m-1} |\ell s + x\rangle \,,$$

where x is some $x \in \{0, \ldots, s-1\}$ with $f(x) = y$, and m is either $\lfloor N/s \rfloor$ or $\lfloor N/s \rfloor + 1$. Applying F then results in

$$\begin{aligned}
\frac{1}{\sqrt{Nm}} \sum_{\ell=0}^{m-1} \sum_{k=0}^{N-1} \omega_N^{(\ell s + x)k} |k\rangle &= \frac{1}{\sqrt{Nm}} \sum_{k=0}^{N-1} \sum_{\ell=0}^{m-1} \omega_N^{\ell s k} \omega_N^{xk} |k\rangle \\
&= \frac{1}{\sqrt{Nm}} \sum_{k=0}^{N-1} \Big(\sum_{\ell=0}^{m-1} \omega_m^{\ell s k m/N} \Big) \omega_N^{xk} |k\rangle \\
&= \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} \hat{1}_m\Big(\frac{ks}{N} m\Big)\, \omega_N^{xk} |k\rangle \,.
\end{aligned}$$

where we used that $\omega_N^t = e^{2\pi i t/N} = e^{2\pi i (t/N) m/m} = \omega_m^{tm/N}$, and where we consider $\hat{1}_m$ as a function on $\mathbb{R}/m\mathbb{Z}$. Recall that in Section 5.2, in the proof of Proposition 5.3, the only k's for which $\hat{1}_m$ did not vanish where the multiples of N/s; therefore, we would observe such a k with certainty. Here, in order to obtain information on s, we may hope that we still get sufficient contribution from those k's that are *close* to a multiple of N/s, so that we observe such a k with sufficient probability upon measuring the register. Indeed, for k of the form $k = cN/s + \delta$ with $|\delta| \leq 1/2$, since $\hat{1}_m$ is periodic we can re-write

$$\hat{1}_m\Big(\frac{ks}{N} m\Big) = \hat{1}_m\Big(cm + \frac{\delta s}{N} m\Big) = \hat{1}_m\Big(\frac{\delta s}{N} m\Big) \,.$$

Furthermore, since $|\delta| \leq 1/2$ and $m \approx N/s$, we expect the argument $\xi := \delta s m/N$ to be bounded in absolute value by $|\xi| \lesssim 1/2$, so that Lemma 5.1 applies and ensures that this particular k will be observed with good probability. Even though this intuition can be rigorously worked out, we provide here a slightly different analysis. Below, $\lfloor \cdot \rceil$ denotes rounding to the nearest integer.

Theorem 5.10. *Let $f : \mathbb{Z} \to \mathcal{R}$ be so that there exists a positive integer $s \leq N = 2^n$ such that $f(x) = f(x')$ if and only if $x - x'$ is an integer multiple of s. Then, the circuit in Figure 5.2*

produces $k \in \{0, \ldots, N-1\}$ of the form $k = \lfloor \ell N/s \rceil$ with probability at least $4/\pi^2$ for a random (but unknown) $\ell \in \{0, \ldots, s-1\}$.

Proof. Consider the subspace spanned by the vectors $|f(x)\rangle$ for $x \in \{0, \ldots, s-1\}$, and let $\tilde{F}$ be the quantum Fourier transform with respect to this basis. Then, we can rewrite the intermediary state of the quantum circuit, after U_f is applied, as

$$\frac{1}{\sqrt{N}} \sum_{x=0}^{N-1} |x\rangle|f(x)\rangle = \frac{1}{\sqrt{N}} \sum_{x=0}^{N-1} |x\rangle \otimes \tilde{F}\tilde{F}^\dagger |f(x)\rangle = \frac{1}{\sqrt{sN}} \sum_{\ell=0}^{s-1} \sum_{x=0}^{N-1} \omega_s^{-x\ell} |x\rangle \otimes \tilde{F}|f(\ell)\rangle \,.$$

Again, for the purpose of the analysis, we assume that the second register is measured, but now in the basis given by the $\tilde{F}|f(\ell)\rangle$'s. As a result, the state collapses to

$$\frac{1}{\sqrt{N}} \sum_{x=0}^{N-1} \omega_s^{-x\ell} |x\rangle$$

for a random $\ell \in \{0, \ldots, s-1\}$. We emphasize that this measurement is a thought experiment, which does not affect the (marginal) distribution of k. Applying F then results in

$$\frac{1}{N} \sum_{x=0}^{N-1} \sum_{k=0}^{N-1} \omega_s^{-x\ell} \omega_N^{xk} |k\rangle = \frac{1}{N} \sum_{k=0}^{N-1} \sum_{x=0}^{N-1} \omega_N^{x(k-\ell N/s)} |k\rangle = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} \hat{1}_N(k - \ell N/s) |k\rangle \,,$$

and therefore, by Lemma 5.1, the final measurement produces $k = \lfloor \ell N/s \rceil$ with probability

$$p_{k|\ell} := \frac{\hat{1}_N(k - \ell N/s)^2}{N} \geq \frac{4}{\pi^2} \,,$$

which proves the claim. □

Given such a k as promised (with constant probability) by Theorem 5.10, we still need to solve the problem of finding s. Choosing $N \geq s^2$, we have that

$$\left| \frac{k}{N} - \frac{\ell}{s} \right| \leq \frac{1}{2N} \leq \frac{1}{2s^2} \,.$$

Thus, we may understand ℓ/s to be a *good* approximation of k/N by means of a fraction with a *small* denominator. Furthermore, it is easy to see that such an approximation is unique. Finally, Theorem 5.14 below shows that such an approximation ℓ/s, with a relation between approximation quality and denominator size as promised here, can be efficiently computed by means of the *continued fraction* of k/N.

Thus, if ℓ and s happen to be coprime (and k is as promised), s can directly be recovered; otherwise, $s' = s/\gcd(\ell, s)$ is obtained. Thus, similarly to period finding in $\mathbb{Z}/N\mathbb{Z}$, we can recover s from two executions *if* the respective ℓ's are coprime and the k's are of the right form $k = \lfloor \ell N/s \rceil$. With the bound/approximation on the probability of two numbers being coprime mentioned in Section 5.2, and the guarantee from Theorem 5.10, the probability of the above happening is shown to be lower bounded by a constant in the range 5% to 10%.

5.7 Continued Fractions

We briefly introduce here the basics of continued fractions, so as to understand the above claim on the efficient computability of good approximations.

Definition 5.4. *For any integer $n \geq 0$, and for any sequence of integers $a_0, a_1, \ldots, a_n$ with $a_0 \geq 0$ and $a_1, \ldots, a_n > 0$, we define $[a_0; a_1, \ldots, a_n]$ to be the* **continued fraction***, recursively defined via $[a_0] := a_0$ and*

$$[a_i; a_{i+1}, \ldots, a_n] := a_i + \frac{1}{[a_{i+1}; a_{i+2}, \ldots, a_n]} .$$

In other words,

$$[a_0; a_1, \ldots, a_n] := a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{\ddots}}}$$

The a_i's are called the **partial quotients***, and $[a_0; a_1, \ldots, a_i]$ is called the i-th* **convergent***.*

We remark that, by definition, we require the partial quotients a_i to be (non-negative) integers; in some cases though, we will explicitly allow a_n to be in $\mathbb{R}$. This then allows us to write

$$[a_0; a_1, \ldots, a_n] = \left[a_0; a_1, \ldots, a_{n-1} + \tfrac{1}{a_n}\right]$$

or, more generally,

$$[a_0; a_1, \ldots, a_n] = \left[a_0; a_1, \ldots, a_{i-1}, [a_i; a_{i+1}, \ldots, a_n]\right] .$$

Proposition 5.11. *Every positive $x \in \mathbb{Q}$ is a (finite) continued fraction $x = [a_0; a_1, \ldots, a_n]$, and the partial quotients are unique up to $[a_0; a_1, \ldots, a_{n-1}, 1] = [a_0; a_1, \ldots, a_{n-1}+1]$.*

Proof. We first note that for $x = [a_0; a_1, \ldots, a_n]$ with $n \geq 1$, $[a_1; a_2, \ldots, a_n] > 1$ unless $n = 1$ and $a_1 = 1$, and thus $x = [x - 1; 1] \in \mathbb{Z}$. This in particular shows uniqueness of a_0 for $x \notin \mathbb{Z}$.

Now, for a given positive $x \in \mathbb{Q}$, write $x = p/q$ for positive coprime integers p, q. If $q = 1$ then $x \in \mathbb{Z}$, and existence is thus clear: $x = [x]$, and (almost) uniqueness follows from the observation that if $x = [a_0; a_1, \ldots, a_n]$ then, by the above, x can be an integer only if $x = [x - 1; 1]$. Otherwise, we can write $p = a_0 q + r$ for integers a_0 and r with $0 < r < q$, and therefore

$$x = \frac{a_0 q + r}{q} = a_0 + \frac{r}{q} = a_0 + \frac{1}{x'}$$

for $x' := q/r > 1$. Given that $r < q$, we may assume by induction that $x' = [a_1; a_2, \ldots, a_n]$ exists and is (almost) unique, and we conclude existence and (almost) uniqueness of $x = [a_0; a_1, \ldots, a_n]$, where the latter exploits uniqueness of a_0 here. □

The above existence proof also shows that the partial quotients can be efficiently computed, essentially by means of the Euclidean algorithm.

Theorem 5.12. *For any continued fraction $x = [a_0; a_1, \ldots, a_n]$, let $(p_{-1}, q_{-1}), (p_0, q_0), \ldots, (p_n, q_n)$ be the sequence recursively defined as*

$$\begin{array}{ll} p_{-1} := 1 \\ q_{-1} := 0 \end{array} , \quad \begin{array}{ll} p_0 := a_0 \\ q_0 := 1 \end{array} \quad \text{and} \quad \begin{array}{ll} p_i := a_i p_{i-1} + p_{i-2} \\ q_i := a_i q_{i-1} + q_{i-2} \end{array}$$

for $i = 1, \ldots, n$. Then

$$\frac{p_i}{q_i} = [a_0; a_1, \ldots, a_i] , \qquad \frac{p_{i-1}}{q_{i-1}} - \frac{p_i}{q_i} = \frac{(-1)^i}{q_i q_{i-1}} \qquad \text{and} \qquad \left|\frac{p_{i-1}}{q_{i-1}} - x\right| \leq \frac{1}{q_i q_{i-1}}$$

for all $i \in \{0, 1, \ldots, n\}$, respectively $i \in \{1, \ldots, n\}$ for the latter two claims.

From the second claim, it follows in particular that for *odd* indices i, the sequence of convergents p_i/q_i is monotonically *decreasing*, and monotonically *increasing* for *even* indices i; see the proof for some more details. Furthermore, given that it implies that $q_i p_{i-1} - p_i q_{i-1} = \pm 1$, the second claim ensures that p_i and q_i are coprime.

We also observe that, given that the q_i's are lower bounded by the Fibonacci numbers, which grow exponentially, the third claim implies that the convergents $[a_0; a_1, \dots, a_i]$ converge fast.

Before proving Theorem 5.12, we first prove the following.

Lemma 5.13. *For $n \geq 1$, let $x = [a_0; a_1, \dots, a_{n-1}, \xi_n]$ with an arbitrary $\xi_n \in \mathbb{R}$. Then,*

$$x = \frac{p_{n-1}\xi_n + p_{n-2}}{q_{n-1}\xi_n + q_{n-2}}\,.$$

Proof. For $n = 1$, the right hand side equals

$$\frac{a_0\xi_1 + 1}{\xi_1} = a_0 + \frac{1}{\xi_1} = x\,.$$

For $n > 1$ we recall that $[a_0; a_1, \dots, a_{n-1}, \xi_n] = [a_0; a_1, \dots, a_{n-1} + \frac{1}{\xi_n}]$ and apply induction to conclude that

$$x = \frac{p_{n-2}\big(a_{n-1} + \frac{1}{\xi_n}\big) + p_{n-3}}{q_{n-2}\big(a_{n-1} + \frac{1}{\xi_n}\big) + q_{n-3}} = \frac{p_{n-2}a_{n-1}\xi_n + p_{n-2} + p_{n-3}\xi_n}{q_{n-2}a_{n-1}\xi_n + q_{n-2} + q_{n-3}\xi_n} = \frac{p_{n-1}\xi_n + p_{n-2}}{q_{n-1}\xi_n + q_{n-2}}\,.$$

Thus, the claim holds for all $n \geq 1$. □

Proof (of Theorem 5.12). For the first claim, we simply apply Lemma 5.13 and conclude that

$$[a_0; a_1, \dots, a_i] = \frac{p_{i-1}a_i + p_{i-2}}{q_{i-1}a_i + q_{i-2}} = \frac{p_i}{q_i}$$

by definition of p_i and q_i. For the second claim, we observe that

$$q_i p_{i-1} - p_i q_{i-1} = (a_i q_{i-1} + q_{i-2})p_{i-1} - (a_i p_{i-1} + p_{i-2})q_{i-1} = p_{i-1}q_{i-2} - q_{i-1}p_{i-2}$$

and conclude by induction. Finally, as for the last claim, we show the more general claim

$$0 \leq (-1)^i \left(\frac{p_{i-1}}{q_{i-1}} - \frac{p_j}{q_j}\right) \leq \frac{1}{q_i q_{i-1}}$$

for all $i \leq j \leq n$. To prove these bounds, we observe that

$$(-1)^i \left(\frac{p_{i-1}}{q_{i-1}} - \frac{p_j}{q_j}\right) = (-1)^i \left(\frac{p_{i-1}}{q_{i-1}} - \frac{p_i}{q_i}\right) - (-1)^{i+1}\left(\frac{p_i}{q_i} - \frac{p_j}{q_j}\right) = \frac{1}{q_i q_{i-1}} - (-1)^{i+1}\left(\frac{p_i}{q_i} - \frac{p_j}{q_j}\right)$$

and so the claimed bounds follow from the second claim in case $j = i$, and for $j > i$ by induction, noting that $q_{i+1} \geq q_{i-1}$. □

Theorem 5.12 in particular shows that the convergents are *good approximations*. What we need in Section 5.6 above is the converse: any good enough approximation must be a convergent.

Theorem 5.14. *If*

$$\left|\frac{p}{q} - x\right| \leq \frac{1}{2q^2}$$

then p/q is a convergent of x.

Proof. We may assume p and q to be coprime. Write $p/q = [a_0; a_1, \ldots, a_n]$ and let $\xi_n \in \mathbb{Q}$ be so that $x = [a_0; a_1, \ldots, a_n, \xi_{n+1}]$. We want to show that $\xi_n \geq 1$; this then implies that we can write $\xi_{n+1} = [b_0; b_1, \ldots, b_m]$ with $b_0 \geq 1$, so that $x = [a_0; a_1, \ldots, a_n, b_0, \ldots, b_m]$ and thus p/q is indeed a convergent of x.

For this purpose, write

$$\frac{p}{q} - x = \varepsilon \frac{\gamma}{q^2}$$

with $\varepsilon = \pm 1$ and $0 \leq \gamma \leq 1/2$. By the degree of freedom in choosing n one more or less, we may assume that $\varepsilon = (-1)^{n-1}$. Using that $p = p_n$ and $q = q_n$, which holds since $p/q = p_n/q_n$ and both are in lowest terms, and applying Lemma 5.13, we obtain

$$\varepsilon \frac{\gamma}{q_n^2} = \frac{p_n}{q_n} - x = \frac{p_n}{q_n} - \frac{p_n \xi_{n+1} + p_{n-1}}{q_n \xi_{n+1} + q_{n-1}} = \frac{p_n q_{n-1} - q_n p_{n-1}}{q_n (q_n \xi_{n+1} + q_{n-1})} = \frac{(-1)^{n-1}}{q_n (q_n \xi_{n+1} + q_{n-1})} \,,$$

where the last equality is due to Theorem 5.12. Thus, $\gamma(q_n \xi_{n+1} + q_{n-1}) = q_n$, and solving for ξ_{n+1} shows that

$$\xi_{n+1} = \frac{q_n - \gamma q_{n-1}}{\gamma q_n} = \frac{1}{\gamma} - \frac{q_{n-1}}{q_n} \geq 1 \,,$$

which concludes the proof. □

5.8 Shor's Factoring Algorithm

We are now ready to understand Shor's quantum algorithm for factoring integers. As a matter of fact, all that is left to do is to to reduce factoring to period finding in $\mathbb{Z}$, and then we can apply the quantum algorithm from Section 5.6 to solve the latter.

Let N be the (odd) integer that we want to factor, and let $N = p_1^{e_1} \cdots p_m^{e_m}$ be its prime factorization. Let a be an integer in the range $\{1, \ldots, N-1\}$ that is coprime to N; if a is not coprime then we immediately can get a nontrivial factor of N. Shor's algorithm now simply works by computing the order $s := \mathrm{ord}(a)$ of a modulo N, i.e., the order of a as element in the group $(\mathbb{Z}/N\mathbb{Z})^*$, by applying the period finding algorithm to the function

$$f : \mathbb{Z} \to (\mathbb{Z}/N\mathbb{Z})^*,\ x \mapsto a^x \,,$$

where we understand a as an element in $(\mathbb{Z}/N\mathbb{Z})^*$. Indeed, by basic algebra, $a^x = a^{x'}$ if and only if $x - x'$ is a multiple of $\mathrm{ord}(a)$. The reduction from factoring to period finding in $\mathbb{Z}$ now follows from the following two propositions.

Proposition 5.15. *Let $N > 2$ be a positive integer and $a \in (\mathbb{Z}/N\mathbb{Z})^*$. If $s := \mathrm{ord}(a)$ is even and $b := a^{s/2}$ is neither 1 nor -1, then $\gcd(b+1, N)$ is a nontrivial divisor of N.*

Proof. Since, as integer, $b^2 \equiv 1 \pmod{N}$, it follows that N divides $b^2 - 1 = (b+1)(b-1)$. However, since $b \not\equiv \pm 1 \pmod{N}$, it follows that N divides neither $b+1$ nor $b-1$. Thus, some non-trivial factor of N must divide $b+1$ but not $b-1$, and vice versa. □

We also offer the following alternative proof (assuming N to be odd), which requires some basic understanding of the structure of $(\mathbb{Z}/N\mathbb{Z})^*$, but prepares for the proof of Proposition 5.16 below. Concretely, the proof makes use of the *Chinese Remainder Theorem*, which states that the canonical map

$$(\mathbb{Z}/N\mathbb{Z})^* \to (\mathbb{Z}/p_1^{e_1}\mathbb{Z})^* \times \cdots \times (\mathbb{Z}/p_m^{e_m}\mathbb{Z})^* \,,\ x \mapsto (x_1, \ldots, x_m) \,,$$

where x_i is the reduction of x modulo $p_1^{e_1}$, is an isomorphism, and of the fact that $(\mathbb{Z}/p^e\mathbb{Z})^*$ is cyclic with even order $p^{e-1}(p-1)$ for odd primes p, and so -1 is the only element with order 2.

Proof. We have $b^2 = a^s = 1$. Thus, for $i \in \{1, \ldots, m\}$, $b_i^2 = 1$ and hence $b_i = \pm 1$. However, as $b \neq \pm 1$, we must have that $b_i = -1$ for some i and $b_j = 1$ for some j. This then implies that, as integer, $b + 1$ is a multiple of $p_i^{e_i}$ but not of $p_j^{e_j}$. □

The following shows that if a is chosen uniformly at random then it satisfies the required properties with probability at least $1/2$, unless N is prime.

Proposition 5.16. *Let $N > 2$ be an odd integer with m distinct prime factors. Then, at most a 2^{-m+1}-fraction of the elements $a \in (\mathbb{Z}/N\mathbb{Z})^*$ are such that $s := \mathrm{ord}(a)$ is odd or $a^{s/2} = \pm 1$.*

Proof. We exploit the above isomorphism of $(\mathbb{Z}/N\mathbb{Z})^*$ into m cyclic groups $(\mathbb{Z}/p_i^{e_i}\mathbb{Z})^*$ with even order. For instance, for any $a \in (\mathbb{Z}/N\mathbb{Z})^*$ the order s of a is the *lcm* of the respective orders s_i of the components a_i. We write $s = 2^d t$ for $0 \leq d \in \mathbb{Z}$ and $1 \leq t \in \mathbb{Z}$ odd, and similarly $s_i = 2^{d_i} t_i$. Then, $d = \max(d_1, \ldots, d_m)$, and thus $d = d_i$ for some i, and $t = lcm(t_1, \ldots, t_m)$.

Assume that $d_j < d$ for some j. Then, in particular, $d \geq 1$ and thus s is even. Furthermore,

$$a_i^{s/2} = a_i^{2^{d-1}t} = \left(a_i^{2^{d_i-1}t_i}\right)^{t/t_i} = \left(a_i^{s_i/2}\right)^{t/t_i} = (-1)^{t/t_i} = -1\,,$$

while, given that $s/2 = 2^{d-1}t$ is a multiple of $s_j = 2^{d_j}t_j$,

$$a_j^{s/2} = 1\,.$$

Thus, $a^{s/2} \neq \pm 1$.

Now, exploiting that $(\mathbb{Z}/p_i^{e_i}\mathbb{Z})^*$ is cyclic with even order, say order $2^{\ell_i} r_i$ with $1 \leq \ell_i \in \mathbb{Z}$ and $1 \leq r_i \in \mathbb{Z}$ odd, half of the elements $a_i \in (\mathbb{Z}/p_i^{e_i}\mathbb{Z})^*$ have $d_i = \ell_i$ (namely all the odd powers of a given generator), while all the other half has $d_i < \ell_i$. Thus, for any i, no matter what $a_1, \ldots, a_i$ and thus $d_1, \ldots, d_i$ are, at most half of the elements a_{i+1} have $d_{i+1} = d_i$, and thus may potentially give rise to an a with s odd or $a^{s/2} = \pm 1$. Thus, at most a $1/2^{m-1}$ fraction of $(a_1, \ldots, a_m)$ may give rise to such an a. □

5.9 Shor's Discrete-Logarithm Algorithm

Let G be a finite cyclic group with generator g. The order q of G may or may not be known. The **discrete logarithm** (with respect to g) of an element $h \in G$ is the unique $s \in \{0, \ldots, q-1\}$ with $g^s = h$. Given such an h, we observe that

$$f : \mathbb{Z}/q\mathbb{Z} \times \mathbb{Z}/q\mathbb{Z} \to G,\ (x, y) \mapsto g^x h^{-y} = g^{x-sy}$$

is such that $f(x, y) = f(x', y')$ if and only if $x - x' = s(y - y')$. Thus, the problem of computing the discrete logarithm reduces to period finding as considered in Section 5.3, up to some insignificant difference in the problem statement. Thus, one is tempted to invoke the quantum circuit from Figure 5.3 to find s. However, there are two issues with this approach: q may not be known, and, even if it is known, it may not be a power of 2, and so it is not clear whether the quantum Fourier transform over $\mathbb{Z}/q\mathbb{Z}$ could be done efficiently.

The solution (to both problems) is to choose a prime power $N = 2^n$ sufficiently large, understand f as a function $\{0, \ldots, N-1\} \times \{0, \ldots, N-1\} \to G$ in the obvious way, and apply the quantum circuit from Figure 5.3, but now with the quantum Fourier transform over $\mathbb{Z}/N\mathbb{Z}$. The analysis is very similar to the analysis of the quantum circuit for period finding in

$\mathbb{Z}$. Indeed, applying F to the first two registers of $|0\rangle|0\rangle|0\rangle$, which are n-qubits each, and then applying U_f, results in

$$|0\rangle|0\rangle|0\rangle \mapsto \frac{1}{N}\sum_{x,y=0}^{N-1}|x\rangle|y\rangle|0\rangle \mapsto \frac{1}{N}\sum_{x,y=0}^{N-1}|x\rangle|y\rangle|f(x,y)\rangle = \frac{1}{N}\sum_{x,y=0}^{N-1}|x\rangle|y\rangle|f(x-sy,0)\rangle\,.$$

Somewhat similar to Section 5.6, we consider the subspace spanned by the vectors $|f(x,0)\rangle$ for $x \in \{0,\ldots,q-1\}$, and we let $\tilde{F}$ be the quantum Fourier transform with respect to this basis. The above intermediary state, after U_f is applied, can then be written as

$$\frac{1}{N}\sum_{x,y=0}^{N-1}|x\rangle|y\rangle|f(x-sy,0)\rangle = \frac{1}{N\sqrt{q}}\sum_{\ell=0}^{q-1}\sum_{x,y=0}^{N-1}\omega_q^{-\ell(x-sy)}|x\rangle|y\rangle\otimes\tilde{F}|f(\ell,0)\rangle\,,$$

which, upon measuring the third register, collapses to

$$\frac{1}{N}\sum_{x,y=0}^{N-1}\omega_q^{-\ell(x-sy)}|x\rangle|y\rangle = \frac{1}{\sqrt{N}}\sum_{x=0}^{N-1}\omega_q^{-\ell x}|x\rangle\otimes\frac{1}{\sqrt{N}}\sum_{y=0}^{N-1}\omega_q^{\ell sy}|y\rangle$$

for some $\ell \in \{0,\ldots,q-1\}$. Applying F to the remaining two registers results in the tensor product of

$$\frac{1}{N}\sum_{x,k}\omega_q^{-\ell x}\omega_N^{xk}|k\rangle = \frac{1}{\sqrt{N}}\sum_{k}\hat{1}_N(k-\ell N/q)|k\rangle$$

and

$$\frac{1}{N}\sum_{y,m}\omega_q^{\ell sy}\omega_N^{ym}|m\rangle = \frac{1}{\sqrt{N}}\sum_{m}\hat{1}_N(m+\ell sN/q)|m\rangle\,.$$

Thus, when measuring, we observe $k = \lfloor \ell N/q \rceil$ and $m = -\lfloor \ell sN/q \rceil$ with probability at least $16/\pi^4$, which is approximately 0.16. By means of the continued fraction techniques, i.e., Theorem 5.14, ℓ and s can then be efficiently recoved.