VII

Quantum Algorithms

Part VII: Contents

23 Quantum states and measurements

In this chapter we'll explain how to set up quantum states using linear algebra. This will allow me to talk about quantum *circuits* in the next chapter, which will set the stage for Shor's algorithm.

I won't do very much physics (read: none at all). That is, I'll only state what the physical reality is in terms of linear algebras, and defer the philosophy of why this is true to your neighborhood "Philosophy of Quantum Mechanics" class (which is a "social science" class at MIT!).

§23.1 Bra-ket notation

Physicists have their own notation for vectors: whereas I previously used something like v, e_1, and so on, in this chapter you'll see the infamous **bra-ket** notation: a vector will be denoted by $|\bullet\rangle$, where $\bullet$ is some variable name: unlike in math or Python, this can include numbers, symbols, Unicode characters, whatever you like. This is called a "ket". To pay a homage to physicists everywhere, we'll use this notation for this chapter too.

Abuse of Notation 23.1.1 (For this part, $\dim H < \infty$)**.** In this part on quantum computation, we'll use the word "Hilbert space" as defined earlier, but in fact all our Hilbert spaces will be finite-dimensional.

If $\dim H = n$, then its orthonormal basis elements are often denoted

$$|0\rangle, |1\rangle, \dots, |n-1\rangle$$

(instead of e_i) and a generic element of H denoted by

$$|\psi\rangle, |\phi\rangle, \dots$$

and various other Greek letters.

Now for any $|\psi\rangle \in H$, we can consider the canonical dual element in $H^\vee$ (since H has an inner form), which we denote by $\langle\psi|$ (a "bra"). For example, if $\dim H = 2$ then we can write

$$|\psi\rangle = \begin{bmatrix} \alpha \\ \beta \end{bmatrix}$$

in an orthonormal basis, in which case

$$\langle\psi| = \begin{bmatrix} \overline{\alpha} & \overline{\beta} \end{bmatrix}.$$

We even can write dot products succinctly in this notation: if $|\phi\rangle = \begin{bmatrix} \gamma \\ \delta \end{bmatrix}$, then the dot product of $\langle\psi|$ and $|\phi\rangle$ is given by

$$\langle\psi|\phi\rangle = \begin{bmatrix} \overline{\alpha} & \overline{\beta} \end{bmatrix} \begin{bmatrix} \gamma \\ \delta \end{bmatrix} = \overline{\alpha}\gamma + \overline{\beta}\delta.$$

So we will use the notation $\langle\psi|\phi\rangle$ instead of the more mathematical $\langle|\psi\rangle, |\phi\rangle\rangle$. In particular, the squared norm of $|\psi\rangle$ is just $\langle\psi|\psi\rangle$. Concretely, for $\dim H = 2$ we have $\langle\psi|\psi\rangle = |\alpha|^2 + |\beta|^2$.

§23.2 The state space

If you think that's weird, well, it gets worse.

In classical computation, a bit is either 0 or 1. More generally, we can think of a classical space of n possible states $0, \dots, n-1$. Thus in the classical situation, the space of possible states is just a discrete set with n elements.

In quantum computation, a **qubit** is instead any *complex linear combination* of 0 and 1. To be precise, consider the normed complex vector space

$$H = \mathbb{C}^{\oplus 2}$$

and denote the orthonormal basis elements by $|0\rangle$ and $|1\rangle$. Then a *qubit* is a nonzero element $|\psi\rangle \in H$, so that it can be written in the form

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

where α and β are not both zero. Typically, we normalize so that $|\psi\rangle$ has norm 1:

$$\langle \psi | \psi \rangle = 1 \iff |\alpha|^2 + |\beta|^2 = 1.$$

In particular, we can recover the "classical" situation with $|0\rangle \in H$ and $|1\rangle \in H$, but now we have some "intermediate" states, such as

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

Philosophically, what has happened is that:

> **Instead of allowing just the states $|0\rangle$ and $|1\rangle$, we allow any complex linear combination of them.**

More generally, if $\dim H = n$, then the possible states are nonzero elements

$$c_0 |0\rangle + c_1 |1\rangle + \dots + c_{n-1} |n-1\rangle$$

which we usually normalize so that $|c_0|^2 + |c_1|^2 + \dots + |c_{n-1}|^2 = 1$.

§23.3 Observations

Prototypical example for this section: id *corresponds to not making a measurement since all its eigenvalues are equal, but any operator with distinct eigenvalues will cause collapse.*

If you think that's weird, well, it gets worse. First, some linear algebra review (Definition 15.4.1):

Definition 23.3.1. Let V be a finite-dimensional inner product space. For a map $T \colon V \to V$, the following conditions are equivalent:

- $\langle Tx, y \rangle = \langle x, Ty \rangle$ for any $x, y \in V$.
- $T = T^\dagger$.

A map T satisfying these conditions is called **Hermitian**.

Question 23.3.2. Show that T is normal.

Thus, we know that T is diagonalizable with respect to the inner form, so for a suitable basis we can write it in an orthonormal basis as

$$T = \begin{bmatrix} \lambda_0 & 0 & \dots & 0 \\ 0 & \lambda_1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_{n-1} \end{bmatrix}.$$

As we've said, this is fantastic: not only do we have a basis of eigenvectors, but the eigenvectors are pairwise orthogonal, and so they form an orthonormal basis of V.

Question 23.3.3. Show that all eigenvalues of T are real. ($T = T^\dagger$.)

Back to quantum computation. Suppose we have a state $|\psi\rangle \in H$, where $\dim H = 2$; we haven't distinguished a particular basis yet, so we just have a nonzero vector. Then the way observations work (and this is physics, so you'll have to take my word for it) is as follows:

Pick a Hermitian operator $T\colon H \to H$; then observations of T return eigenvalues of T.

To be precise:

- Pick a Hermitian operator $T\colon H \to H$, which is called the **observable**.
- Consider its eigenvalues $\lambda_0, \dots, \lambda_{n-1}$ and corresponding eigenvectors $|0\rangle_T, \dots, |n-1\rangle_T$. Tacitly we may assume that $|0\rangle_T, \dots, |n-1\rangle_T$ form an orthonormal basis of H. (The subscript T is here to distinguish the eigenvectors of T from the basis elements of H.)
- Write $|\psi\rangle$ in the orthonormal basis as

$$c_0 |0\rangle_T + c_1 |1\rangle_T + \cdots + c_{n-1} |n-1\rangle_T .$$

- Then the probability of observing λ_i is

$$\frac{|c_i|^2}{|c_0|^2 + \cdots + |c_{n-1}|^2}.$$

  This is called making an **observation along T**.

Note that in particular, for any nonzero constant c, $|\psi\rangle$ and $c|\psi\rangle$ are indistinguishable, which is why we like to normalize $|\psi\rangle$. But the queerest thing of all is what happens to $|\psi\rangle$: by measuring it, we actually destroy information. This behavior is called **quantum collapse**.

- Suppose for simplicity that we observe $|\psi\rangle$ with T and obtain an eigenvalue λ, and that $|i\rangle_T$ is the only eigenvector with this eigenvalue. Then, the state $|\psi\rangle$ *collapses* to just the state $c_i |i\rangle_T$: all the other information is destroyed. (In fact, we may as well say it collapses to $|i\rangle_T$, since again constant factors are not relevant.)

- More generally, if we observe λ, consider the generalized eigenspace H_λ (i.e. the span of eigenvectors with the same eigenvalue). Then the physical state $|\psi\rangle$ has been changed as well: it has now been projected onto the eigenspace H_λ. In still other words, after observation, the state collapses to
$$\sum_{\substack{0 \le i \le n \\ \lambda_i = \lambda}} c_i |i\rangle_T .$$

In other words,

When we make a measurement, the coefficients from different eigenspaces are destroyed.

Why does this happen? Beats me... physics (and hence real life) is weird. But anyways, an example.

Example 23.3.4 (Quantum measurement of a state $|\psi\rangle$)
Let $H = \mathbb{C}^{\oplus 2}$ with orthonormal basis $|0\rangle$ and $|1\rangle$ and consider the state
$$|\psi\rangle = \frac{i}{\sqrt{5}}|0\rangle + \frac{2}{\sqrt{5}}|1\rangle = \begin{bmatrix} i/\sqrt{5} \\ 2/\sqrt{5} \end{bmatrix} \in H.$$

(a) Let
$$T = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}.$$
This has eigenvectors $|0\rangle = |0\rangle_T$ and $|1\rangle = |1\rangle_T$, with eigenvalues $+1$ and -1. So if we measure $|\psi\rangle$ to T, we get $+1$ with probability $1/5$ and -1 with probability $4/5$. After this measurement, the original state collapses to $|0\rangle$ if we measured $+1$, and $|1\rangle$ if we measured -1. So we never learn the original probabilities.

(b) Now consider $T = \mathrm{id}$, and arbitrarily pick two orthonormal eigenvectors $|0\rangle_T$, $|1\rangle_T$; thus $\psi = c_0|0\rangle_T + c_1|1\rangle_T$. Since all eigenvalues of T are $+1$, our measurement will always be $+1$ no matter what we do. But there is also no collapsing, because none of the coefficients get destroyed.

(c) Now consider
$$T = \begin{bmatrix} 0 & 7 \\ 7 & 0 \end{bmatrix}.$$
The two normalized eigenvectors are
$$|0\rangle_T = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ 1 \end{bmatrix} \qquad |1\rangle_T = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ -1 \end{bmatrix}$$
with eigenvalues $+7$ and -7 respectively. In this basis, we have
$$|\psi\rangle = \frac{2+i}{\sqrt{10}}|0\rangle_T + \frac{-2+i}{\sqrt{10}}|1\rangle_T .$$

So we get $+7$ with probability $\frac{1}{2}$ and -7 with probability $\frac{1}{2}$, and after the measurement, $|\psi\rangle$ collapses to one of $|0\rangle_T$ and $|1\rangle_T$.

Question 23.3.5. Suppose we measure $|\psi\rangle$ with T and get λ. What happens if we measure with T again?

For $H = \mathbb{C}^{\oplus 2}$ we can come up with more classes of examples using the so-called **Pauli matrices**. These are the three Hermitian matrices

$$\sigma_z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \qquad \sigma_x = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \qquad \sigma_y = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}.$$

These matrices are important because:

Question 23.3.6. Show that these three matrices, plus the identity matrix, form a basis for the set of Hermitian 2×2 matrices.

So the Pauli matrices are a natural choice of basis.[1]

Their normalized eigenvectors are

$$|\uparrow\rangle = |0\rangle = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \qquad |\downarrow\rangle = |1\rangle = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

$$|\rightarrow\rangle = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \qquad |\leftarrow\rangle = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

$$|\otimes\rangle = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ i \end{bmatrix} \qquad |\odot\rangle = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -i \end{bmatrix}$$

which we call "z-up", "z-down", "x-up", "x-down", "y-up", "y-down" respectively. (The eigenvalues are $+1$ for "up" and -1 for "down".) So, given a state $|\psi\rangle \in \mathbb{C}^{\oplus 2}$ we can make a measurement with respect to any of these three bases by using the corresponding Pauli matrix.

In light of this, the previous examples were (a) measuring along σ_z, (b) measuring along id, and (c) measuring along $7\sigma_x$.

Notice that if we are given a state $|\psi\rangle$, and are told in advance that it is either $|\rightarrow\rangle$ or $|\leftarrow\rangle$ (or any other orthogonal states) then we are in what is more or less a classical situation. Specifically, if we make a measurement along σ_x, then we find out which state that $|\psi\rangle$ was in (with 100% certainty), and the state does not undergo any collapse. Thus, orthogonal states are reliably distinguishable.

§23.4 Entanglement

Prototypical example for this section: Singlet state: spooky action at a distance.

If you think that's weird, well, it gets worse.

Qubits don't just act independently: they can talk to each other by means of a *tensor product*. Explicitly, consider

$$H = \mathbb{C}^{\oplus 2} \otimes \mathbb{C}^{\oplus 2}$$

[1] Well, natural due to physics reasons.

endowed with the norm described in Problem 13D*. One should think of this as a qubit A in a space H_A along with a second qubit B in a different space H_B, which have been allowed to interact in some way, and $H = H_A \otimes H_B$ is the set of possible states of *both* qubits. Thus

$$|0\rangle_A \otimes |0\rangle_B, \quad |0\rangle_A \otimes |1\rangle_B, \quad |1\rangle_A \otimes |0\rangle_B, \quad |1\rangle_A \otimes |1\rangle_B$$

is an orthonormal basis of H; here $|i\rangle_A$ is the basis of the first $\mathbb{C}^{\oplus 2}$ while $|i\rangle_B$ is the basis of the second $\mathbb{C}^{\oplus 2}$, so these vectors should be thought of as "unrelated" just as with any tensor product. The pure tensors mean exactly what you want: for example $|0\rangle_A \otimes |1\rangle_B$ means "0 for qubit A and 1 for qubit B".

As before, a measurement of a state in H requires a Hermitian map $H \to H$. In particular, if we only want to measure the qubit B along M_B, we can use the operator

$$\mathrm{id}_A \otimes M_B.$$

The eigenvalues of this operator coincide with the ones for M_B, and the eigenspace for λ will be the $H_A \otimes (H_B)_\lambda$, so when we take the projection the A qubit will be unaffected.

This does what you would hope for pure tensors in H:

Example 23.4.1 (Two non-entangled qubits)

Suppose we have qubit A in the state $\frac{i}{\sqrt{5}}|0\rangle_A + \frac{2}{\sqrt{5}}|1\rangle_A$ and qubit B in the state $\frac{1}{\sqrt{2}}|0\rangle_B + \frac{1}{\sqrt{2}}|1\rangle_B$. So, the two qubits in tandem are represented by the pure tensor

$$|\psi\rangle = \left(\frac{i}{\sqrt{5}}|0\rangle_A + \frac{2}{\sqrt{5}}|1\rangle_A\right) \otimes \left(\frac{1}{\sqrt{2}}|0\rangle_B + \frac{1}{\sqrt{2}}|1\rangle_B\right).$$

Suppose we measure $|\psi\rangle$ along

$$M = \mathrm{id}_A \otimes \sigma_z^B.$$

The eigenspace decomposition is

- $+1$ for the span of $|0\rangle_A \otimes |0\rangle_B$ and $|1\rangle_A \otimes |0\rangle_B$, and
- -1 for the span of $|0\rangle_A \otimes |1\rangle_B$ and $|1\rangle_A \otimes |1\rangle_B$.

(We could have used other bases, like $|{\to}\rangle_A \otimes |0\rangle_B$ and $|{\leftarrow}\rangle_A \otimes |0\rangle_B$ for the first eigenspace, but it doesn't matter.) Expanding $|\psi\rangle$ in the four-element basis, we find that we'll get the first eigenspace with probability

$$\left|\frac{i}{\sqrt{10}}\right|^2 + \left|\frac{2}{\sqrt{10}}\right|^2 = \frac{1}{2}.$$

and the second eigenspace with probability $\frac{1}{2}$ as well. (Note how the coefficients for A don't do anything!) After the measurement, we destroy the coefficients of the other eigenspace; thus (after re-normalization) we obtain the collapsed state

$$\left(\frac{i}{\sqrt{5}}|0\rangle_A + \frac{2}{\sqrt{5}}|1\rangle_A\right) \otimes |0\rangle_B \qquad \text{or} \qquad \left(\frac{i}{\sqrt{5}}|0\rangle_A + \frac{2}{\sqrt{5}}|1\rangle_A\right) \otimes |1\rangle_B$$

again with 50% probability each.

So this model lets us more or less work with the two qubits independently: when we make the measurement, we just make sure to not touch the other qubit (which corresponds to the identity operator).

Exercise 23.4.2. Show that if $\mathrm{id}_A \otimes \sigma_x^B$ is applied to the $|\psi\rangle$ in this example, there is no collapse at all. What's the result of this measurement?

Since the $\otimes$ is getting cumbersome to write, we say:

Abuse of Notation 23.4.3. From now on $|0\rangle_A \otimes |0\rangle_B$ will be abbreviated to just $|00\rangle$, and similarly for $|01\rangle$, $|10\rangle$, $|11\rangle$.

Example 23.4.4 (Simultaneously measuring a general 2-Qubit state)
Consider a normalized state $|\psi\rangle$ in $H = \mathbb{C}^{\oplus 2} \otimes \mathbb{C}^{\oplus 2}$, say

$$|\psi\rangle = \alpha\,|00\rangle + \beta\,|01\rangle + \gamma\,|10\rangle + \delta\,|11\rangle\,.$$

We can make a measurement along the diagonal matrix $T\colon H \to H$ with

$$T(|00\rangle) = 0\,|00\rangle\,, \quad T(|01\rangle) = 1\,|01\rangle\,, \quad T(|10\rangle) = 2\,|10\rangle\,, \quad T(|11\rangle) = 3\,|11\rangle\,.$$

Thus we get each of the eigenvalues 0, 1, 2, 3 with probability $|\alpha|^2$, $|\beta|^2$, $|\gamma|^2$, $|\delta|^2$. So if we like we can make "simultaneous" measurements on two qubits in the same way that we make measurements on one qubit.

However, some states behave very weirdly.

Example 23.4.5 (The singlet state)
Consider the state

$$|\Psi_-\rangle = \frac{1}{\sqrt{2}}\,|01\rangle - \frac{1}{\sqrt{2}}\,|10\rangle$$

which is called the **singlet state**. One can see that $|\Psi_-\rangle$ is not a simple tensor, which means that it doesn't just consist of two qubits side by side: the qubits in H_A and H_B have become *entangled*.

Now, what happens if we measure just the qubit A? This corresponds to making the measurement

$$T = \sigma_z^A \otimes \mathrm{id}_B.$$

The eigenspace decomposition of T can be described as:

- The span of $|00\rangle$ and $|01\rangle$, with eigenvalue $+1$.
- The span of $|10\rangle$ and $|11\rangle$, with eigenvalue -1.

So one of two things will happen:

- With probability $\frac{1}{2}$, we measure $+1$ and the collapsed state is $|01\rangle$.
- With probability $\frac{1}{2}$, we measure -1 and the collapsed state is $|10\rangle$.

But now we see that measurement along A has told us what the state of the bit B is completely!

By solely looking at measurements on A, we learn B; this paradox is called *spooky action at a distance*, or in Einstein's tongue, **spukhafte Fernwirkung**. Thus,

In tensor products of Hilbert spaces, states which are not pure tensors correspond to "entangled" states.

What this really means is that the qubits cannot be described independently; the state of the system must be given as a whole. That's what entangled states mean: the qubits somehow depend on each other.

§23.5 A few harder problems to think about

Problem 23A. We measure $|\Psi_-\rangle$ by $\sigma_x^A \otimes \mathrm{id}_B$, and hence obtain either $+1$ or -1. Determine the state of qubit B from this measurement.

Problem 23B (Greenberger-Horne-Zeilinger paradox)**.** Consider the state in $(\mathbb{C}^{\oplus 2})^{\otimes 3}$

$$|\Psi\rangle_{\text{GHZ}} = \frac{1}{\sqrt{2}} \left(|0\rangle_A |0\rangle_B |0\rangle_C - |1\rangle_A |1\rangle_B |1\rangle_C \right).$$

Find the value of the measurements along each of

$$\sigma_y^A \otimes \sigma_y^B \otimes \sigma_x^C, \quad \sigma_y^A \otimes \sigma_x^B \otimes \sigma_y^C, \quad \sigma_x^A \otimes \sigma_y^B \otimes \sigma_y^C, \quad \sigma_x^A \otimes \sigma_x^B \otimes \sigma_x^C.$$

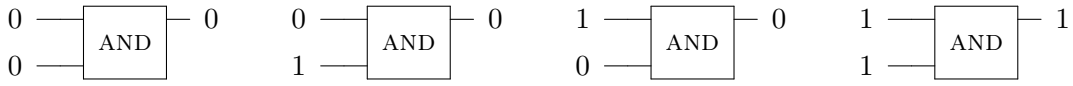
As for the paradox: what happens if you multiply all these measurements together?

24 Quantum circuits

Now that we've discussed qubits, we can talk about how to use them in circuits. The key change — and the reason that quantum circuits can do things that classical circuits cannot — is the fact that we are allowing linear combinations of 0 and 1.

§24.1 Classical logic gates

In classical logic, we build circuits which take in some bits for input, and output some more bits for input. These circuits are built out of individual logic gates. For example, the **AND gate** can be pictured as follows.



One can also represent the AND gate using the "truth table":

A	B	A and B
0	0	0
0	1	0
1	0	0
1	1	1

Similarly, we have the **OR gate** and the **NOT gate**:

A	B	A or B
0	0	0
0	1	1
1	0	1
1	1	1

A	not A
0	1
1	0

We also have a so-called **COPY gate**, which duplicates a bit.

Of course, the first theorem you learn about these gates is that:

Theorem 24.1.1 (AND, OR, NOT, COPY are universal)
The set of four gates AND, OR, NOT, COPY is universal in the sense that any boolean function $f\colon \{0,1\}^n \to \{0,1\}$ can be implemented as a circuit using only these gates.

Proof. Somewhat silly: we essentially write down a circuit that OR's across all input strings in $f^{\text{pre}}(1)$. For example, suppose we have $n = 3$ and want to simulate the function $f(abc)$ with $f(011) = f(110) = 1$ and 0 otherwise. Then the corresponding Boolean expression for f is simply

$$f(abc) = [(\text{not } a) \text{ and } b \text{ and } c] \text{ or } [a \text{ and } b \text{ and } (\text{not } c)].$$

Clearly, one can do the same for any other f, and implement this logic into a circuit. □

Remark 24.1.2 — Since x and $y = \text{not } ((\text{not } x) \text{ or } (\text{not } y))$, it follows that in fact, we can dispense with the AND gate.

§24.2 Reversible classical logic

Prototypical example for this section: CNOT gate, Toffoli gate.

For the purposes of quantum mechanics, this is not enough. To carry through the analogy we in fact need gates that are **reversible**, meaning the gates are bijections from the input space to the output space. In particular, such gates must take the same number of input and output gates.

Example 24.2.1 (Reversible gates)

(a) None of the gates AND, OR, COPY are reversible for dimension reasons.

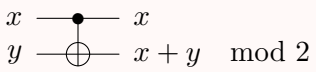
(b) The NOT gate, however, is reversible: it is a bijection $\{0,1\} \to \{0,1\}$.

Example 24.2.2 (The CNOT gate)

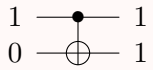
The controlled-NOT gate, or the **CNOT** gate, is a reversible 2-bit gate with the following truth table.

In		Out	
0	0	0	0
1	0	1	1
0	1	0	1
1	1	1	0

In other words, this gate XOR's the first bit to the second bit, while leaving the first bit unchanged. It is depicted as follows.



The first dot is called the "control", while the $\oplus$ is the "negation" operation: the first bit controls whether the second bit gets flipped or not. Thus, a typical application might be as follows.



So, NOT and CNOT are the only nontrivial reversible gates on two bits.

We now need a different definition of universal for our reversible gates.

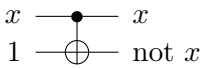
Definition 24.2.3. A set of reversible gates can **simulate** a Boolean function $f(x_1 \dots x_n)$, if one can implement a circuit which takes

- As input, $x_1 \dots x_n$ plus some fixed bits set to 0 or 1, called **ancilla bits**[1].
- As output, the input bits $x_1, \dots, x_n$, the output bit $f(x_1, \dots, x_n)$, and possibly some extra bits (called **garbage bits**).

[1]The English word "ancilla" means "maid".

The gate(s) are **universal** if they can simulate any Boolean function.

For example, the CNOT gate can simulate the NOT gate, using a single ancilla bit 1, according to the following circuit.



Unfortunately, it is not universal.

Proposition 24.2.4 (CNOT $\not\Rightarrow$ AND)

The CNOT gate cannot simulate the boolean function "x and y".

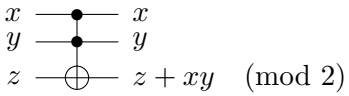
Sketch of Proof. One can see that any function simulated using only CNOT gates must be of the form

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n \pmod 2$$

because CNOT is the map $(x, y) \mapsto (x, x + y)$. Thus, even with ancilla bits, we can only create functions of the form $ax + by + c \pmod 2$ for fixed a, b, c. The AND gate is not of this form. □

So, we need at least a three-qubit gate. The most commonly used one is:

Definition 24.2.5. The three-bit **Toffoli gate**, also called the CCNOT gate, is given by



So the Toffoli has two controls, and toggles the last bit if and only if both of the control bits are 1.

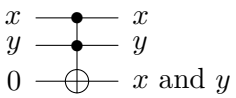
This replacement is sufficient.

Theorem 24.2.6 (Toffoli gate is universal)

The Toffoli gate is universal.

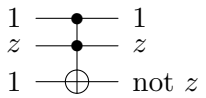
Proof. We will show it can *reversibly* simulate AND, NOT, hence OR, which we know is enough to show universality. (We don't need COPY because of reversibility.)

For the AND gate, we draw the circuit



with one ancilla bit, and no garbage bits.

For the NOT gate, we use two ancilla 1 bits and one garbage bit:



This completes the proof. □

Hence, in theory we can create any classical circuit we desire using the Toffoli gate alone. Of course, this could require exponentially many gates for even the simplest of functions. Fortunately, this is NO BIG DEAL because I'm a math major, and having 2^n gates is a problem best left for the CS majors.

§24.3 Quantum logic gates

In quantum mechanics, since we can have *linear combinations* of basis elements, our logic gates will instead consist of *linear maps*. Moreover, in quantum computation, gates are always reversible, which was why we took the time in the previous section to show that we can still simulate any function when restricted to reversible gates (e.g. using the Toffoli gate).

First, some linear algebra:

Definition 24.3.1. Let V be a finite dimensional inner product space. Then for a map $U : V \to V$, the following are equivalent:

- $\langle U(x), U(y) \rangle = \langle x, y \rangle$ for $x, y \in V$.
- $U^\dagger$ is the inverse of U.
- $\|x\| = \|U(x)\|$ for $x \in V$.

The map U is called **unitary** if it satisfies these equivalent conditions.

Then

Quantum logic gates are unitary matrices.

In particular, unlike the classical situation, quantum gates are always reversible (and hence they always take the same number of input and output bits).

For example, consider the CNOT gate. Its quantum analog should be a unitary map $U_{\text{CNOT}} : H \to H$, where $H = \mathbb{C}^{\oplus 2} \otimes \mathbb{C}^{\oplus 2}$, given on basis elements by

$$U_{\text{CNOT}}(|00\rangle) = |00\rangle, \quad U_{\text{CNOT}}(|01\rangle) = |01\rangle$$

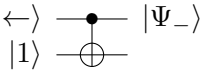
$$U_{\text{CNOT}}(|10\rangle) = |11\rangle, \quad U_{\text{CNOT}}(|11\rangle) = |10\rangle.$$

So pictorially, the quantum CNOT gate is given by

OK, so what? The whole point of quantum mechanics is that we allow linear qubits to be in linear combinations of $|0\rangle$ and $|1\rangle$, too, and this will produce interesting results. For example, let's take $|\leftarrow\rangle = \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle)$ and plug it into the top, with $|1\rangle$ on the bottom, and see what happens:

$$U_{\text{CNOT}}\left(|\leftarrow\rangle \otimes |1\rangle\right) = U_{\text{CNOT}}\left(\frac{1}{\sqrt{2}}(|01\rangle - |11\rangle)\right) = \frac{1}{\sqrt{2}}\left(|01\rangle - |10\rangle\right) = |\Psi_-\rangle$$

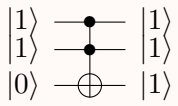
which is the fully entangled *singlet state*! Picture:



Thus, when we input mixed states into our quantum gates, the outputs are often entangled states, even when the original inputs are not entangled.

Example 24.3.2 (More examples of quantum gates)

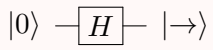
(a) Every reversible classical gate that we encountered before has a quantum analog obtained in the same way as CNOT: by specifying the values on basis elements. For example, there is a quantum Tofolli gate which for example sends



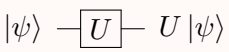
(b) The **Hadamard gate** on one qubit is a rotation given by

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

Thus, it sends $|0\rangle$ to $|\rightarrow\rangle$ and $|1\rangle$ to $|\leftarrow\rangle$. Note that the Hadamard gate is its own inverse. It is depicted by an "H" box.



(c) More generally, if U is a 2×2 unitary matrix (i.e. a map $\mathbb{C}^{\oplus 2} \to \mathbb{C}^{\oplus 2}$) then there is **$U$-rotation gate** similar to the previous one, which applies U to the input.



For example, the classical NOT gate is represented by $U = \sigma_x$.

(d) A **controlled U-rotation gate** generalizes the CNOT gate. Let $U\colon \mathbb{C}^{\oplus 2} \to \mathbb{C}^{\oplus 2}$ be a rotation gate, and let $H = \mathbb{C}^{\oplus 2} \otimes \mathbb{C}^{\oplus 2}$ be a 2-qubit space. Then the controlled U gate has the following circuit diagrams.

Thus, U is applied when the controlling bit is 1, and CNOT is the special case $U = \sigma_x$. As before, we get interesting behavior if the control is mixed.

And now, some more counterintuitive quantum behavior. Suppose we try to use CNOT as a copy, with truth table.

In		Out	
0	0	0	0
1	0	1	1
0	1	0	1
1	1	1	0

The point of this gate is to be used with a garbage 0 at the bottom to try and simulate a "copy" operation. So indeed, one can check that

Thus we can copy $|0\rangle$ and $|1\rangle$. But as we've already seen if we input $|\leftarrow\rangle \otimes |0\rangle$ into U, we end up with the entangled state $|\Psi_-\rangle$ which is decisively *not* the $|\leftarrow\rangle \otimes |\leftarrow\rangle$ we wanted.

And in fact, the so-called **no-cloning theorem** implies that it's impossible to duplicate an arbitrary $|\psi\rangle$; the best we can do is copy specific orthogonal states as in the classical case. See also Problem 24B.

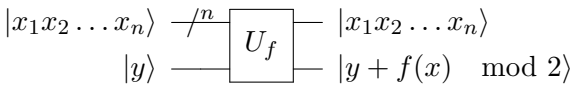
§24.4 Deutsch-Jozsa algorithm

The Deutsch-Jozsa algorithm is the first example of a nontrivial quantum algorithm which cannot be performed classically: it is a "proof of concept" that would later inspire Grover's search algorithm and Shor's factoring algorithm.

The problem is as follows: we're given a function $f\colon \{0,1\}^n \to \{0,1\}$, and promised that the function f is either

- A constant function, or
- A balanced function, meaning that exactly half the inputs map to 0 and half the inputs map to 1.

The function f is given in the form of a reversible black box U_f which is the control of a NOT gate, so it can be represented as the circuit diagram



i.e. if $f(x_1,\ldots,x_n)=0$ then the gate does nothing, otherwise the gate flips the y bit at the bottom. The slash with the n indicates that the top of the input really consists of n qubits, not just the one qubit drawn, and so the black box U_f is a map on $n+1$ qubits.

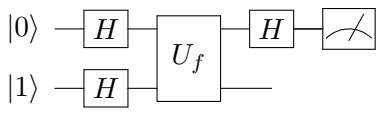
The problem is to determine, with as few calls to the black box U_f as possible, whether f is balanced or constant.

> **Question 24.4.1.** Classically, show that in the worst case we may need up to $2^{n-1}+1$ calls to the function f to answer the question.

So with only classical tools, it would take $O(2^n)$ queries to determine whether f is balanced or constant. However,

> **Theorem 24.4.2** (Deutsch-Jozsa)
> The Deutsch-Jozsa problem can be determined in a quantum circuit with only a single call to the black box.

Proof. For concreteness, we do the case $n=1$ explicitly; the general case is contained in Problem 24C. We claim that the necessary circuit is



Here the H's are Hadamard gates, and the meter at the end of the rightmost wire indicates that we make a measurement along the usual $|0\rangle$, $|1\rangle$ basis. This is not a typo! Even though classically the top wire is just a repeat of the input information, we are about to see that it's the top we want to measure.

Note that after the two Hadamard operations, the state we get is

$$\begin{aligned}|01\rangle &\xmapsto{H^{\otimes 2}} \left(\frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)\right) \otimes \left(\frac{1}{\sqrt{2}}(|0\rangle - |1\rangle)\right) \\ &= \frac{1}{2}\Big(|0\rangle \otimes (\,|0\rangle - |1\rangle\,) \;+\; |1\rangle \otimes (\,|0\rangle - |1\rangle\,)\Big).\end{aligned}$$

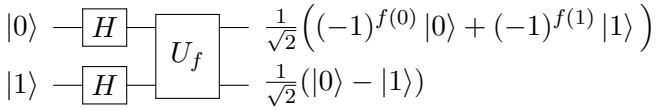
So after applying U_f, we obtain

$$\frac{1}{2}\Big(|0\rangle \otimes (\,|0 + f(0)\rangle - |1 + f(0)\rangle\,) \;+\; |1\rangle \otimes (\,|0 + f(1)\rangle - |1 + f(1)\rangle\,)\Big)$$

where the modulo 2 has been left implicit. Now, observe that the effect of going from $|0\rangle - |1\rangle$ to $|0 + f(x)\rangle - |1 + f(x)\rangle$ is merely to either keep the state the same (if $f(x) = 0$) or to negate it (if $f(x) = 1$). So we can simplify and factor to get

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

Thus, the picture so far is:



In particular, the resulting state is not entangled, and we can simply discard the last qubit (!). Now observe:

- If f is constant, then the upper-most state is $\pm|{\rightarrow}\rangle$.
- If f is balanced, then the upper-most state is $\pm|{\leftarrow}\rangle$.

So simply doing a measurement along σ_x will give us the answer. Equivalently, perform another H gate (so that $H|{\rightarrow}\rangle = |0\rangle$, $H|{\leftarrow}\rangle = |1\rangle$) and measuring along σ_z in the usual $|0\rangle$, $|1\rangle$ basis. Thus for $n = 1$ we only need a single call to the oracle. □

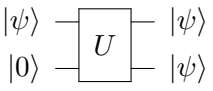
§24.5 A few harder problems to think about

Problem 24A (Fredkin gate)**.** The **Fredkin gate** (also called the controlled swap, or CSWAP gate) is the three-bit gate with the following truth table:

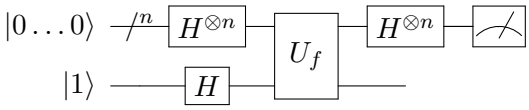
In			Out		
0	0	0	0	0	0
0	0	1	0	0	1
0	1	0	0	1	0
0	1	1	0	1	1
1	0	0	1	0	0
1	0	1	1	1	0
1	1	0	1	0	1
1	1	1	1	1	1

Thus the gate swaps the last two input bits whenever the first bit is 1. Show that this gate is also reversible and universal.

Problem 24B (Baby no-cloning theorem)**.** Show that there is no unitary map U on two qubits which sends $U(|\psi\rangle \otimes |0\rangle) = |\psi\rangle \otimes |\psi\rangle$ for any qubit $|\psi\rangle$, i.e. the following circuit diagram is impossible.



Problem 24C (Deutsch-Jozsa)**.** Given the black box U_f described in the Deutsch-Jozsa algorithm, consider the following circuit.



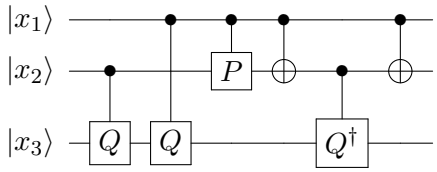
That is, take n copies of $|0\rangle$, apply the Hadamard rotation to all of them, apply U_f, reverse the Hadamard to all n input bits (again discarding the last bit), then measure all n bits in the $|0\rangle/|1\rangle$ basis (as in Example 23.4.4).

Show that the probability of measuring $|0\dots0\rangle$ is 1 if f is constant and 0 if f is balanced.

Problem 24D† (Barenco et al, 1995; arXiv:quant-ph/9503016v1)**.** Let

$$P = \begin{bmatrix} 1 & 0 \\ 0 & i \end{bmatrix} \qquad Q = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & -i \\ -i & 1 \end{bmatrix}$$

Verify that the quantum Toffoli gate can be implemented using just controlled rotations via the circuit



This was a big surprise to researchers when discovered, because classical reversible logic requires three-bit gates (e.g. Toffoli, Fredkind).

25 Shor's algorithm

OK, now for Shor's Algorithm: how to factor $M = pq$ in $O\left((\log M)^2\right)$ time.

This is arguably the reason agencies such as the US's National Security Agency have been diverting millions of dollars toward quantum computing.

§25.1 The classical (inverse) Fourier transform

The "crux move" in Shor's algorithm is the so-called quantum Fourier transform. The Fourier transform is used to extract *periodicity* in data, and it turns out the quantum analogue is a lot faster than the classical one.

Let me throw the definition at you first. Let N be a positive integer, and let $\omega_N = \exp\left(\frac{2\pi i}{N}\right)$.

Definition 25.1.1. Given a tuple of complex numbers

$$(x_0, x_1, \dots, x_{N-1})$$

its **discrete inverse Fourier transform** is the sequence $(y_0, y_1, \dots, y_{N-1})$ defined by

$$y_k = \frac{1}{N} \sum_{j=0}^{N-1} \omega_N^{jk} x_j.$$

Equivalently, one is applying the matrix

$$\frac{1}{N} \begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ 1 & \omega_N & \omega_N^2 & \dots & \omega_N^{N-1} \\ 1 & \omega_N^2 & \omega_N^4 & \dots & \omega_N^{2(N-1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \omega_N^{N-1} & \omega_N^{2(N-1)} & \dots & \omega_N^{(N-1)^2} \end{bmatrix} \begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_{N-1} \end{bmatrix} = \begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_{N-1} \end{bmatrix}.$$

The reason this operation is important is because it lets us detect if the x_i are periodic. More generally, given a sequence of 1's appearing with period r, the amplitudes will peak at inputs which are divisible by $\frac{N}{\gcd(N,r)}$. Mathematically, we have that

$$x_k = \sum_{j=0}^{N-1} y_j \omega_N^{-jk}.$$

Example 25.1.2 (Example of discrete inverse Fourier transform)
Let $N = 6$, $\omega = \omega_6 = \exp(\frac{2\pi i}{6})$ and suppose $(x_0, x_1, x_2, x_3, x_4, x_5) = (0, 1, 0, 1, 0, 1)$

(hence x_i is periodic modulo 2). Thus,

$$\begin{aligned} y_0 &= \tfrac{1}{6}\left(\omega^0 + \omega^0 + \omega^0\right) = 1/2 \\ y_1 &= \tfrac{1}{6}\left(\omega^1 + \omega^3 + \omega^5\right) = 0 \\ y_2 &= \tfrac{1}{6}\left(\omega^2 + \omega^6 + \omega^{10}\right) = 0 \\ y_3 &= \tfrac{1}{6}\left(\omega^3 + \omega^9 + \omega^{15}\right) = -1/2 \\ y_4 &= \tfrac{1}{6}\left(\omega^4 + \omega^{12} + \omega^{20}\right) = 0 \\ y_5 &= \tfrac{1}{6}\left(\omega^5 + \omega^{15} + \omega^{25}\right) = 0. \end{aligned}$$

Thus, in the inverse transformation the "amplitudes" are all concentrated at multiples of 3; thus this reveals the periodicity of the original sequence by $\frac{N}{3} = 2$.

Remark 25.1.3 — The fact that this operation is called the "inverse" Fourier transform is mostly a historical accident (as my understanding goes). Confusingly, the corresponding quantum operation is the (not-inverted) Fourier transform.

If we apply the definition as written, computing the transform takes $O(N^2)$ time. It turns out that by a classical algorithm called the **fast Fourier transform** (whose details we won't discuss, but it effectively "reuses" calculations), one can reduce this to $O(N \log N)$ time. However, for Shor's algorithm this is also insufficient; we need something like $O\left((\log N)^2\right)$ instead. This is where the quantum Fourier transform comes in.

§25.2 The quantum Fourier transform

Note that to compute a Fourier transform, we need to multiply an $N \times N$ matrix with an N-vector, so this takes $O(N^2)$ multiplications. However, we are about to show that with a quantum computer, one can do this using $O((\log N)^2)$ quantum gates when $N = 2^n$, on a system with n qubits.

First, some more notation:

Abuse of Notation 25.2.1. In what follows, $|x\rangle$ will refer to $|x_n\rangle \otimes |x_{n-1}\rangle \otimes \cdots \otimes |x_1\rangle$ where $x = x_n x_{n-1} \ldots x_1$ in binary. For example, if $n = 3$ then $|6\rangle$ really means $|1\rangle\otimes|1\rangle\otimes|0\rangle$. Likewise, we refer to $0.x_1x_2\ldots x_n$ as binary.

Observe that the n-qubit space now has an orthonormal basis $|0\rangle$, $|1\rangle$, …, $|N-1\rangle$

Definition 25.2.2. Consider an n-qubit state

$$|\psi\rangle = \sum_{k=0}^{N-1} x_k |k\rangle .$$

The **quantum Fourier transform** is defined by

$$U_{\text{QFT}}(|\psi\rangle) = \frac{1}{\sqrt{N}} \sum_{j=0}^{N-1} \left(\sum_{k=0}^{N-1} \omega_N^{jk} x_k \right) |j\rangle .$$

In other words, using the basis $|0\rangle, \dots, |N-1\rangle$, U_{QFT} is given by the matrix

$$U_{\text{QFT}} = \frac{1}{\sqrt{N}} \begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ 1 & \omega_N & \omega_N^2 & \dots & \omega_N^{N-1} \\ 1 & \omega_N^2 & \omega_N^4 & \dots & \omega_N^{2(N-1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \omega_N^{N-1} & \omega_N^{2(N-1)} & \dots & \omega_N^{(N-1)^2} \end{bmatrix}$$

This is the exactly the same definition as before, except we have a $\sqrt{N}$ factor added so that U_{QFT} is unitary. But the trick is that in the quantum setup, the matrix can be rewritten:

Proposition 25.2.3 (Tensor representation)
Let $|x\rangle = |x_n x_{n-1} \dots x_1\rangle$. Then

$$\begin{aligned} U_{\text{QFT}}(|x_n x_{n-1} \dots x_1\rangle) = \frac{1}{\sqrt{N}} &(|0\rangle + \exp(2\pi i \cdot 0.x_1)\,|1\rangle) \\ &\otimes (|0\rangle + \exp(2\pi i \cdot 0.x_2 x_1)\,|1\rangle) \\ &\otimes \dots \\ &\otimes (|0\rangle + \exp(2\pi i \cdot 0.x_n \dots x_1)\,|1\rangle) \end{aligned}$$

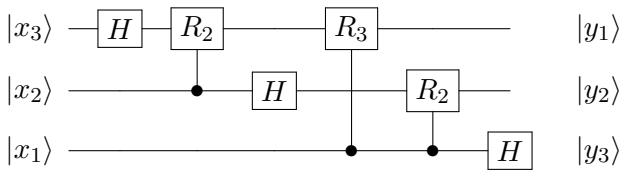
Proof. Direct (and quite annoying) computation. In short, expand everything. □

So by using mixed states, the quantum Fourier transform can use this "multiplication by tensor product" trick that isn't possible classically.

Now, without further ado, here's the circuit. Define the rotation matrices

$$R_k = \begin{bmatrix} 1 & 0 \\ 0 & \exp(2\pi i/2^k) \end{bmatrix}.$$

Then, for $n = 3$ the circuit is given by using controlled R_k's as follows:

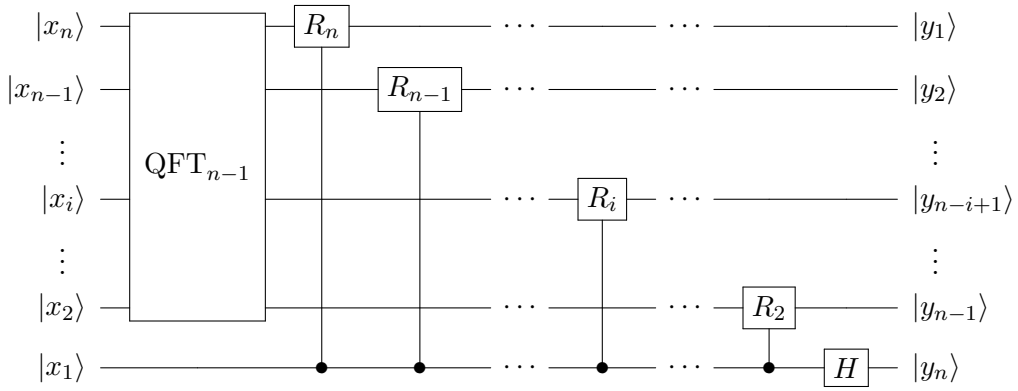


Exercise 25.2.4. Show that in this circuit, the image of $|x_3 x_2 x_1\rangle$ is

$$\Big(|0\rangle + \exp(2\pi i \cdot 0.x_1)\,|1\rangle\Big) \otimes \Big(|0\rangle + \exp(2\pi i \cdot 0.x_2 x_1)\,|1\rangle\Big) \otimes \Big(|0\rangle + \exp(2\pi i \cdot 0.x_3 x_2 x_1)\,|1\rangle\Big)$$

as claimed.

For general n, we can write this as inductively as



Question 25.2.5. Convince yourself that when $n = 3$ the two circuits displayed are equivalent.

Thus, the quantum Fourier transform is achievable with $O(n^2)$ gates, which is enormously better than the $O(N \log N)$ operations achieved by the classical fast Fourier transform (where $N = 2^n$).

§25.3 Shor's algorithm

The quantum Fourier transform is the key piece of Shor's algorithm. Now that we have it, we can solve the factoring problem.

Let $p, q > 3$ be odd primes, and assume $p \neq q$. The main idea is to turn factoring an integer $M = pq$ into a problem about finding the order of $x \pmod M$; the latter is a "periodicity" problem that the quantum Fourier transform will let us solve. Specifically, say that an $x \pmod M$ is *good* if

(i) $\gcd(x, M) = 1$,

(ii) The order r of $x \pmod M$ is even, and

(iii) Factoring $0 \equiv (x^{r/2}-1)(x^{r/2}+1) \pmod M$, neither of the two factors is $0 \pmod M$. Thus one of them is divisible by p, and the other is divisible by q.

Exercise 25.3.1 (For contest number theory practice)**.** Show that for $M = pq$ at least half of the residues in $(\mathbb{Z}/M\mathbb{Z})^\times$ are good.

So if we can find the order of an arbitrary $x \in (\mathbb{Z}/M\mathbb{Z})^\times$, then we just keep picking x until we pick a good one (this happens more than half the time); once we do, we compute $\gcd(x^{r/2} - 1, M)$ using the Euclidean algorithm to extract one of the prime factors of M, and we're home free.

Now how do we do this? The idea is not so difficult: first we generate a sequence which is periodic modulo r.

Example 25.3.2 (Factoring 77: generating the periodic state)
Let's say we're trying to factor $M = 77$, and we randomly select $x = 2$, and want to

find its order r. Let $n = 13$ and $N = 2^{13}$, and start by initializing the state

$$|\psi\rangle = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} |k\rangle .$$

Now, build a circuit U_x (depending on x) which takes $|k\rangle |0\rangle$ to $|k\rangle |x^k \bmod M\rangle$. Applying this to $|\psi\rangle \otimes |0\rangle$ gives

$$U(|\psi\rangle |0\rangle) = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} |k\rangle \otimes |x^k \bmod M\rangle .$$

Now suppose we measure the second qubit, and get a state of $|128\rangle$. That tells us that the collapsed state now, up to scaling, is

$$(|7\rangle + |7+r\rangle + |7+2r\rangle + \dots) \otimes |128\rangle .$$

The bottleneck is actually the circuit U_x; one can compute $x^k \pmod M$ by using repeated squaring, but it's still the clumsy part of the whole operation.

In general, the operation is:

- Pick a sufficiently large $N = 2^n$ (say, $N \geq M^2$).
- Generate $|\psi\rangle = \sum_{k=0}^{2^n-1} |k\rangle$.
- Build a circuit U_x which computes $|x^k \mod M\rangle$.
- Apply it to get a state $\frac{1}{\sqrt{N}} \sum_{k=0}^{2^n-1} |k\rangle \otimes |x^k \mod M\rangle$.
- Measure the second qubit to cause the first qubit to collapse to something which is periodic modulo r. Let $|\phi\rangle$ denote the left qubit.

Suppose we apply the quantum Fourier transform to the left qubit $|\phi\rangle$ now: since the left bit is periodic modulo r, we expect the transform will tell us what r is. Unfortunately, this doesn't quite work out, since N is a power of two, but we don't expect r to be.

Nevertheless, consider a state

$$|\phi\rangle = |k_0\rangle + |k_0 + r\rangle + \dots$$

so for example previously we had $k_0 = 7$ if we measured 128 on $x = 2$. Applying the quantum Fourier transform, we see that the coefficient of $|j\rangle$ in the transformed image is equal to

$$\omega_N^{k_0 j} \cdot \left(\omega_N^0 + \omega_N^{jr} + \omega_N^{2jr} + \omega_N^{3jr} + \dots \right)$$

As this is a sum of roots of unity, we realize we have destructive interference unless $\omega_N^{jr} = 1$ (since N is large). In other words, we approximately have

$$U_{\text{QFT}}(|\phi\rangle) \approx \sum_{\substack{0 \leq j < N \\ jr/N \in \mathbb{Z}}} |j\rangle$$

up to scaling as usual. The bottom line is that

If we measure $U_{\text{QFT}} |\phi\rangle$ we obtain a $|j\rangle$ such that $\frac{jr}{N}$ is close to an $s \in \mathbb{Z}$.

And thus given sufficient luck we can use continued fractions to extract the value of r.

Example 25.3.3 (Finishing the factoring of $M = 77$)

As before, we made an observation to the second qubit, and thus the first qubit collapses to the state $|\phi\rangle = |7\rangle + |7+r\rangle + \dots$. Now we make a measurement and obtain $j = 4642$, which means that for some integer s we have

$$\frac{4642r}{2^{13}} \approx s.$$

Now, we analyze the continued fraction of $\frac{4642}{2^{13}}$; we find the first few convergents are

$$0,\ 1,\ \frac{1}{2},\ \frac{4}{7},\ \frac{13}{23},\ \frac{17}{30},\ \frac{1152}{2033},\ \dots$$

So $\frac{17}{30}$ is a good approximation, hence we deduce $s = 17$ and $r = 30$ as candidates. And indeed, one can check that $r = 30$ is the desired order.

This won't work all the time[1] (for example, we could get unlucky and measure $j = 0$, i.e. $s = 0$, which would tell us no information at all).

But one can show that we succeed any time that

$$\gcd(s, r) = 1.$$

This happens at least $\frac{1}{\log r}$ of the time, and since $r < M$ this means that given sufficiently many trials, we will eventually extract the correct order r. This is Shor's algorithm.

[1] Not to mention the general issue of noise, but that's for engineers to worry about.