

Lecture 2: Quantum bootcamp + Shor's algorithm in detail

Instructor: Fermi Ma

September 21, 2026

Let's start with a quick recap of the steps of Shor's algorithm, which we began in the last lecture:

1. Classical step: reduce factoring to period-finding. This involves picking a random $a \in \{2, 3, \dots, N-1\}$, and showing that if we can find the period of $a^1 \bmod N, a^2 \bmod N, a^3 \bmod N, \dots$, we can solve factoring.
2. Quantum step: use a quantum computer to solve the period-finding problem.

At a high level, the quantum period-finding algorithm works as follows:

1. Pick some Q between N^2 and $2N^2$ such that Q is a power of 2. Picking Q to be a power of 2 will make it simpler to implement the Quantum Fourier Transform, which will come up later.
2. Prepare a uniform superposition of k from 0 to $Q-1$, and then compute the function $f(k) = a^k \bmod N$ in superposition. This yields the state

$$\frac{1}{\sqrt{Q}} \sum_{k=0}^{Q-1} |k\rangle \otimes |a^k \bmod N\rangle. \quad (1)$$

3. Measure the second register, obtaining some $y \in \{0, 1, \dots, N-1\}$. The first register collapses to the superposition of all k such that $a^k \equiv y \bmod N$. If s is the smallest integer such that $a^s \equiv y \bmod N$, and r is the period, then this state has the form (omitting normalization)

$$|s\rangle + |s+r\rangle + |s+2r\rangle + \dots. \quad (2)$$

4. Apply the mod Q Quantum Fourier Transform. The result is a state whose measurement outcomes are concentrated near the values $\frac{\ell Q}{r}$ for $\ell \in \{0, \dots, r-1\}$. If we measure this first register and get the integer z , then we know that $z \approx \frac{\ell Q}{r}$ for some (unknown) integer ℓ .

Here, z and Q are known and we want to solve for ℓ and r . This can be done with a technique called continued fractions, which we will not cover here. But the intuition is basically: because $\frac{z}{Q}$ is an accurate approximation of $\frac{\ell}{r}$ (because Q is chosen to be at least N^2 , and r is at most N), there are only so many possibilities for what r can be. For instance, if I told you that $\frac{6701}{10001} \approx \frac{\ell}{r}$ where r is at most about 100, you might reasonably guess that $\frac{\ell}{r} = \frac{67}{100}$, since $\frac{6701}{10001} = \frac{67}{100}$ exactly.

1 Quantum background

For the rest of this lecture, we'll cover the quantum information / quantum computing background necessary to understand Shor's algorithm.

1.1 What is a quantum state?

The basic unit of information in quantum computation is a qubit, which extends the notion of a classical bit. To represent the state of a qubit, we write

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

where α and β are complex numbers satisfying

$$|\alpha|^2 + |\beta|^2 = 1.$$

The notation $|0\rangle$ and $|1\rangle$ is shorthand for

$$|0\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad |1\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \quad (3)$$

and so

$$\alpha |0\rangle + \beta |1\rangle = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}. \quad (4)$$

If I measure the qubit in the *standard basis*, which refers to the basis $\{|0\rangle, |1\rangle\}$, I'll get 0 with probability $|\alpha|^2$ and 1 with probability $|\beta|^2$. And moreover, measurement changes the state: if my measurement outcome is $b \in \{0, 1\}$, the state after measurement *collapses* to $|b\rangle$.

Physical intuition: Just as a classical bit can describe the state of a light switch, which is either on or off, a qubit can describe the spin of an electron along some direction, where we associate $|0\rangle$ with spin up and $|1\rangle$ with spin down. The key difference between these two settings is that, for a light switch, the classical bit of “on” or “off” fully describes the state of the system, but for a qubit, you need the complex amplitudes α and β in to fully describe the relevant state, e.g., if you want to predict what will happen if you measure the spin in a different direction.

Two-qubit states. What happens when we have more than one qubit? Let's start with two qubits. If the first qubit is in the state

$$|\psi\rangle = a |0\rangle + b |1\rangle \quad (5)$$

and the second qubit is in state

$$|\phi\rangle = c |0\rangle + d |1\rangle, \quad (6)$$

then the joint state of the two-qubit system is described by their **tensor product**:

$$|\psi\rangle \otimes |\phi\rangle = ac |00\rangle + ad |01\rangle + bc |10\rangle + bd |11\rangle. \quad (7)$$

You can verify that this is a valid quantum state: the squared magnitudes of the amplitudes sums to 1. Moreover, the measurement probabilities behave as you expect, e.g., the probability of measuring 00 is exactly $|a|^2|c|^2$.

More generally, a two-qubit state is represented by a vector in $\mathbb{C}^4$ of unit norm:

$$\alpha_{00} |00\rangle + \alpha_{01} |01\rangle + \alpha_{10} |10\rangle + \alpha_{11} |11\rangle. \quad (8)$$

Not every two-qubit state can be written as a tensor product of two one-qubit states. For example, consider the state

$$\frac{1}{\sqrt{2}}(|00\rangle + |11\rangle) = \begin{pmatrix} 1/\sqrt{2} \\ 0 \\ 0 \\ 1/\sqrt{2} \end{pmatrix}. \quad (9)$$

This state cannot be written as the tensor product of two one-qubit states, i.e., there are no complex numbers a, b, c, d satisfying

$$\begin{pmatrix} 1/\sqrt{2} \\ 0 \\ 0 \\ 1/\sqrt{2} \end{pmatrix} = \begin{pmatrix} ac \\ ad \\ bc \\ bd \end{pmatrix}, \quad (10)$$

since multiplying the equations corresponding to the first and fourth entries says that $abcd = 1/2$, but multiplying the equations corresponding to the second and third entries says that $abcd = 0$. State such as this that cannot be written as a tensor product of two states are called **entangled**.

Partial measurement. What happens if I only measure the first qubit of $\frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$? I'll get 0 with probability 1/2 and 1 with probability 1/2. And if I get 0, the whole state collapses to $|00\rangle$ and if I get 1, the whole state collapses to $|11\rangle$. So, if we measure one part of an entangled state and condition on getting a certain outcome, this can change the state of the qubits that we did not measure.

More generally, an n -qubit state has the form $\sum_{x \in \{0,1\}^n} \alpha_x |x\rangle$ where $\sum_x |\alpha_x|^2 = 1$. A full measurement of all n -qubits in the standard basis gives x with probability $|\alpha_x|^2$. If we do a partial measurement of the qubits, say, we measure the first m qubits, then we obtain outcome $w \in \{0,1\}^m$ with probability

$$\sum_{z \in \{0,1\}^{n-m}} |\alpha_{w,z}|^2, \quad (11)$$

which is equal to the probability that the first m qubits are measured to be w if we had simply measured all n qubits. And, conditioned on getting outcome w , the state collapses to the state

$$\frac{1}{\sqrt{\sum_{z \in \{0,1\}^{n-m}} |\alpha_{w,z}|^2}} \sum_{z \in \{0,1\}^{n-m}} \alpha_{w,z} |w, z\rangle \quad (12)$$

$$= |w\rangle \otimes \frac{1}{\sqrt{\sum_{z \in \{0,1\}^{n-m}} |\alpha_{w,z}|^2}} \sum_{z \in \{0,1\}^{n-m}} \alpha_{w,z} |z\rangle, \quad (13)$$

i.e., only the components of the original state consistent with the measurement outcome w survive, and the remaining vector is renormalized to have norm 1.

1.2 Quantum Computation

Okay, so now we know that quantum states are described by complex vectors of unit norm, and we know what happens when you measure them. The next question is: what can a quantum computer do to a quantum state before you measure it?

Quantum computers transform quantum states by applying *unitary transformations* to them. There are many equivalent definitions of a unitary transformation, and it's helpful to be comfortable with all of them.

Definition of a unitary transformation: A matrix $U \in \mathbb{C}^{2^n \times 2^n}$ is a unitary if and only if any of the following equivalent conditions hold:

- The columns $U_0, \dots, U_{2^n-1}$ are an orthonormal basis of $\mathbb{C}^{2^n}$. That is, a unitary is a linear transformation that maps the standard basis $|0\rangle, \dots, |2^n-1\rangle$ to the orthonormal basis $|U_0\rangle, \dots, |U_{2^n-1}\rangle$ where $|U_x\rangle$ is the x -th column of U .
- $U^\dagger \cdot U = \mathbb{I}$, where $U^\dagger$ is the conjugate transpose of U . So $U^\dagger$ is the inverse of U .
- For all unit vectors $|\psi\rangle \in \mathbb{C}^{2^n}$, $U|\psi\rangle$ is a unit vector. That is, applying U preserves the norm.
- For all pairs of vectors $|\psi\rangle, |\phi\rangle \in \mathbb{C}^{2^n}$, $\langle\phi|U^\dagger U|\psi\rangle = \langle\phi|\psi\rangle$, i.e., U preserves inner products. Here $\langle\phi|$ denotes the conjugate transpose of $|\phi\rangle$.

Not every unitary can be implemented *efficiently* on a quantum computer. (Note that even writing down an arbitrary unitary requires specifying exponentially many numbers.) The situation is analogous for classical computation. Classical circuits implement functions, e.g., $f : \{0,1\}^n \mapsto \{0,1\}^m$, but not all functions can be implemented efficiently. Instead, we build classical circuits out of smaller components called gates, and we say that a classical computation is efficient if you can implement with a circuit that has $\text{poly}(n)$ gates from some set of universal gates (e.g., AND/OR/NOT).

Likewise, a quantum circuit is composed of *quantum gates*, where each gate is a unitary transformation acting on a small number of qubits. We say that a unitary transformation is efficient if you can implement it with a circuit that has $\text{poly}(n)$ gates. There are many ways to choose a universal set of quantum gates, from which any unitary can be approximated to arbitrary precision. For today, all we'll need to know is that universal gate sets for quantum computing exist, but the exact choice will not be too important.

There is one particularly important gate that we will need for Shor's algorithm: **the Hadamard gate**. The Hadamard gate is defined as:

$$H = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}, \quad (14)$$

which acts as

$$H|0\rangle = \frac{|0\rangle + |1\rangle}{\sqrt{2}}, \quad H|1\rangle = \frac{|0\rangle - |1\rangle}{\sqrt{2}}. \quad (15)$$

We use the Hadamard gate to the uniform superposition of k over $\{0, 1, \dots, Q-1\}$ in the first step of Shor's algorithm. In particular, since $Q = 2^q$, we start from $|0^q\rangle$, and apply H to each of the q qubits. This gives

$$H|0\rangle \otimes \dots \otimes H|0\rangle = \left(\frac{1}{\sqrt{2}}(|0\rangle + |1\rangle) \right) \otimes \dots \otimes \left(\frac{1}{\sqrt{2}}(|0\rangle + |1\rangle) \right) \quad (16)$$

$$= \frac{1}{\sqrt{2^q}} \sum_{k \in \{0,1\}^q} |k\rangle \quad (17)$$

$$= \frac{1}{\sqrt{Q}} \sum_{k \in \{0,1,\dots,Q-1\}} |k\rangle. \quad (18)$$

Implementing (reversible) classical computation. The next fact that we'll need for Shor's algorithm is that quantum computers can perform efficient classical computation in superposition. As we discussed last time, we use this fact to implement the transformation:

$$\frac{1}{\sqrt{Q}} \sum_{k \in \{0,1,\dots,Q-1\}} |k\rangle \mapsto \frac{1}{\sqrt{Q}} \sum_{k \in \{0,1,\dots,Q-1\}} |k\rangle |a^k \bmod N\rangle. \quad (19)$$

In particular, we're computing the function $f(k) = a^k \bmod N$ **in superposition** (recall that $f(k)$ can be computed efficiently with a classical circuit in $\text{poly}(\log k)$ time using repeated exponentiation).

How do we do this? Note that $k \mapsto a^k \bmod N$ is not a reversible transformation, whereas unitary transformations must be reversible. Thus, to implement a function f , we typically implement the unitary transformation U_f that maps, for all x, y ,

$$U_f |x\rangle |y\rangle = |x\rangle |y \oplus f(x)\rangle. \quad (20)$$

Moreover, as long as you have an efficient classical circuit for f , there is an efficient quantum circuit for U_f .

We're not going to go into details, but you can implement such a transformation as a quantum circuit by taking any classical circuit for computing f and replacing each of its gates with a reversible version, possibly using some extra workspace bits. You then compute the circuit gate by gate, eventually obtaining $y \oplus f(x)$ along with some intermediate "garbage," and finally run the computation backwards to uncompute this garbage.

The result is then, that you can implement a unitary transformation that maps

$$|k\rangle |0\rangle \mapsto |k\rangle |a^k \bmod N\rangle \quad (21)$$

and because this is a linear transformation, doing this on the uniform superposition state gives

$$\frac{1}{\sqrt{Q}} \sum_{k \in \{0,1,\dots,Q-1\}} |k\rangle |0\rangle \mapsto \frac{1}{\sqrt{Q}} \sum_{k \in \{0,1,\dots,Q-1\}} |k\rangle |a^k \bmod N\rangle. \quad (22)$$

1.3 Filling in the rest of Shor's algorithm: the QFT

In the last step of Shor's algorithm we measure the second register and get some value y . The first register collapses to a state of the form

$$|s\rangle + |s+r\rangle + |s+2r\rangle + \dots$$

This state is a superposition of basis states with spacing equal to the period r , but shifted by some unknown s . Let's call this state $|\psi_s\rangle$. In a moment, we'll consider versions of this state for other settings of s , e.g., we'll use $|\psi_0\rangle$ to denote $|0\rangle + |r\rangle + |2r\rangle + \dots$.

Now, if s really were 0, then we would simply measure this state $|\psi_0\rangle$ and learn a random multiple of r (and if we repeated this several times, we could learn r). Unfortunately, since s is a random shift, if we simply measure $|\psi_s\rangle$ in the standard basis, we're going to get a random value and learn nothing about r .

So, a natural question is: can we apply a unitary U (in particular, an efficiently computable one) to $|\psi_s\rangle$, such that measuring $U|\psi_s\rangle$ will allow us to learn r ?

We spoiled the answer earlier: it turns out that setting U to be the Quantum Fourier Transform mod Q will suffice. But suppose we didn't know this. How could we have come up with the idea that U should be the QFT?

Let's first state some desiderata for U . It would be great if, somehow, U has the property that if we apply U to $|\psi_s\rangle$ and measure it, that:

- the measurement outcome is independent of the shift s , but
- the measurement gives us information about r .

There is another useful way to think about this choice of the unitary U . So far, whenever we've talked about measurement, we've measured in the standard basis, i.e., $|0\rangle, \dots, |Q-1\rangle$. But we can also measure in any other orthonormal basis $|\phi_0\rangle, \dots, |\phi_{Q-1}\rangle$, provided we can efficiently implement a unitary U that maps between this basis and the standard basis. That is, if U satisfies

$$U|\phi_z\rangle = |z\rangle \quad (23)$$

for every z , then applying U and measuring in the standard basis is equivalent to measuring in the basis $|\phi_0\rangle, \dots, |\phi_{Q-1}\rangle$. So, choosing a unitary U to apply before measurement is equivalent to choosing a different basis to measure the original state.

So we can rephrase our first item as follows: **we want to find an orthonormal basis in which shifting the state by s does not change the distribution of measurement outcomes.**

For the moment, let's consider the clean case where r divides Q .

Since we're talking about this shift operation, it will be helpful to define the following shift operator:

$$T|x\rangle = |x+1\rangle. \quad (24)$$

where if $x = Q-1$, then $x+1$ wraps around to 0 (i.e., $x+1$ means $x+1 \bmod Q$). Note two things about T . First, T is a unitary transformation. Second,

$$|\psi_s\rangle = T^s |\psi_0\rangle. \quad (25)$$

Now, having defined the shift operator T , our goal is to find an orthonormal basis such that shifting a state does not change its measurement probabilities in this basis. Let's denote these basis states as

$$|\widehat{0}\rangle, |\widehat{1}\rangle, \dots, |\widehat{Q-1}\rangle. \quad (26)$$

If we measure any such basis state $|\widehat{z}\rangle$ in this basis, the result is deterministic: we'll always get the outcome corresponding to $|\widehat{z}\rangle$ with probability 1 (just as measuring a standard basis state $|x\rangle$ in the standard basis always gives the outcome corresponding to $|x\rangle$ with probability 1).

So, if applying T is not supposed to change this measurement outcome, then $T|\widehat{z}\rangle$ must still be the same state $|\widehat{z}\rangle$, up to phase. In other words,

$$T|\widehat{z}\rangle = \lambda|\widehat{z}\rangle, \quad (27)$$

where λ is a complex number with $|\lambda| = 1$.

This means **the basis we're looking for is an orthonormal basis of eigenvectors of the shift operator T .**

Now we can simply solve for these eigenvectors. We want to find all vectors $|\phi\rangle$ with the property that

$$T|\phi\rangle = \lambda|\phi\rangle. \quad (28)$$

Write

$$|\phi\rangle = \sum_{x=0}^{Q-1} \alpha_x |x\rangle. \quad (29)$$

Then

$$\begin{aligned} T|\phi\rangle &= \sum_{x=0}^{Q-1} \alpha_x T|x\rangle && \text{(expanding out } |\phi\rangle\text{)} \\ &= \sum_{x=0}^{Q-1} \alpha_x |x+1 \bmod Q\rangle && \text{(using the definition of } T\text{)} \\ &= \sum_{x=0}^{Q-1} \alpha_{x-1} |x\rangle, && \text{(re-indexing the summation)} \end{aligned}$$

where the indices on the α_x 's are taken mod Q .

Setting this equal to $\lambda|\phi\rangle$ gives

$$\sum_{x=0}^{Q-1} \alpha_{x-1} |x\rangle = \sum_{x=0}^{Q-1} \lambda \alpha_x |x\rangle, \quad (30)$$

and hence, for every x ,

$$\alpha_{x-1} = \lambda \alpha_x. \quad (31)$$

Equivalently,

$$\alpha_x = \lambda^{-1} \alpha_{x-1}, \quad (32)$$

so

$$\alpha_x = \lambda^{-x} \alpha_0. \quad (33)$$

Since the indices are mod Q , we have $\alpha_Q = \alpha_0$. Therefore

$$\lambda^{-Q} \alpha_0 = \alpha_0, \quad (34)$$

and hence

$$\lambda^Q = 1. \quad (35)$$

Thus, λ **must be a Q th root of unity.**

Letting

$$\omega_Q := e^{2\pi i/Q}, \quad (36)$$

the possible values of λ are

$$1, \omega_Q, \omega_Q^2, \dots, \omega_Q^{Q-1}. \quad (37)$$

Taking $\lambda = \omega_Q^z$ and choosing $\alpha_0 = 1/\sqrt{Q}$ gives

$$|\hat{z}\rangle = \frac{1}{\sqrt{Q}} \sum_{x=0}^{Q-1} \omega_Q^{-xz} |x\rangle, \quad (38)$$

for $z \in \{0, 1, \dots, Q-1\}$. You can verify that these states form an orthonormal basis. This is the mod- Q **Fourier basis**.

So, to recap: we were looking for an orthonormal basis such that measurements in this basis are independent of shifts, and we derived the Fourier basis.

Let's quickly verify that the Fourier basis really gives us the first property we wanted. Suppose we write an arbitrary state in the Fourier basis as

$$|\psi\rangle = \sum_{z=0}^{Q-1} \beta_z |\hat{z}\rangle. \quad (39)$$

Since

$$T |\hat{z}\rangle = \omega_Q^z |\hat{z}\rangle, \quad (40)$$

applying T^s gives

$$T^s |\psi\rangle = \sum_{z=0}^{Q-1} \beta_z \omega_Q^{sz} |\hat{z}\rangle. \quad (41)$$

So shifting the state only multiplies each amplitude β_z by a phase. The probability of measuring z is still $|\beta_z|^2$! Thus, measuring in the Fourier basis is agnostic to the unknown shift s .

What about our second requirement? Why does measuring in the Fourier basis tell us anything about r ?

Let QFT_Q denote the unitary that maps the Fourier basis to the standard basis. For all $x \in \{0, 1, \dots, Q\}$, this unitary maps

$$\text{QFT}_Q |x\rangle = \frac{1}{\sqrt{Q}} \sum_{z=0}^{Q-1} \omega_Q^{xz} |z\rangle. \quad (42)$$

Thus, applying QFT_Q and then measuring in the standard basis is exactly the same thing as measuring the original state in the Fourier basis.

Since we're assuming that r divides Q , the normalized state after the partial measurement is

$$|\psi_s\rangle = \sqrt{\frac{r}{Q}} \sum_{j=0}^{Q/r-1} |s + jr\rangle. \quad (43)$$

Applying the QFT gives

$$\begin{aligned}
\text{QFT}_Q |\psi_s\rangle &= \sqrt{\frac{r}{Q}} \sum_{j=0}^{Q/r-1} \text{QFT}_Q |s + jr\rangle && \text{(definition of } |\psi_s\rangle \text{ + linearity)} \\
&= \sqrt{\frac{r}{Q}} \sum_{j=0}^{Q/r-1} \frac{1}{\sqrt{Q}} \sum_{z=0}^{Q-1} \omega_Q^{(s+jr)z} |z\rangle && \text{(using the definition of } \text{QFT}_Q \text{)} \\
&= \frac{\sqrt{r}}{Q} \sum_{j=0}^{Q/r-1} \sum_{z=0}^{Q-1} \omega_Q^{(s+jr)z} |z\rangle && (44) \\
&= \frac{\sqrt{r}}{Q} \sum_{z=0}^{Q-1} \omega_Q^{sz} \left(\sum_{j=0}^{Q/r-1} \omega_Q^{jrz} \right) |z\rangle. && \text{(rearranging the sum)}
\end{aligned}$$

Now what we just need to understand the term:

$$\sum_{j=0}^{Q/r-1} \omega_Q^{jrz}. \quad (45)$$

If z is a multiple of Q/r , then every term in the sum is a Q -th root of unity raised to a multiple of Q , i.e., every time is 1. So the sum is Q/r .

Otherwise, the terms are roots of unity that wrap around the unit circle and cancel to 0. Putting this all together,

$$\text{QFT}_Q |\psi_s\rangle = \frac{1}{\sqrt{r}} \sum_{c=0}^r \omega_Q^{sz} |cQ/r\rangle. \quad (46)$$

Therefore, the only possible measurement outcomes are

$$z \in \left\{ 0, \frac{Q}{r}, \frac{2Q}{r}, \dots, \frac{(r-1)Q}{r} \right\}. \quad (47)$$

So, somewhat magically, measuring in the Fourier basis does exactly the two things we wanted! The unknown shift s only appears in the phase ω_Q^{sz} , so doesn't affect measurement probabilities. At the same time, the period r determines exactly which measurement outcomes show up!

In particular, up to normalization,

$$|0\rangle + |r\rangle + |2r\rangle + \dots \xrightarrow{\text{QFT}_Q} |0\rangle + |Q/r\rangle + |2Q/r\rangle + \dots. \quad (48)$$

And moreover, any shift on the left-hand side shows up on the right-hand side (note: it has to show up *somewhere*, since information can't be lost), but only in the phases.

We technically still haven't covered the case where r *doesn't* divide Q . In a nutshell, the above analysis still approximately holds even when r doesn't divide Q , except that the Fourier-transformed state has support on integers close to Q/r (as Q/r is not an integer). This turns out to be enough to recover r . Please ask an LLM for more details if you're curious :)