Shor's factorization algorithm

The Shor factorization algorithm is the most important application of the quantum Fourier transform and one of the main reasons for the interest in quantum computers. The algorithm would allow probabilistic polynomial factorization of large numbers.

From the number theoretical point of view, this is nothing new: the basis of Shor's algorithm is Fermat's factorization algorithm, in which the factorization of . N is obtained from the knowledge of two numbers a, b, satisfying $a^2 \equiv b^2 \mod N$, thanks to the relation

$$(a+b)(a-b) \equiv 0 \mod N.$$

Fermat's procedure can be used, in particular, if we know some element a and its even order r in the multiplicative group $\mathbb{Z}_N$. Then we have

$$(a^{\frac{r}{2}}+1)(a^{\frac{r}{2}}-1) \equiv 0 \mod N,$$

which provides factorization of N if and only if $a^{\frac{r}{2}}$ is not equal to $-1 \mod N$. Thus, Shor's factorization algorithm for composite odd N looks like this:

- choose $a \in \mathbb{Z}_N^*$ at random (choosing a non-invertible element leads to a factorization immediately)
- find the order r of the element a in $\mathbb{Z}_N^*$
- if r is odd or if $a^{\frac{r}{2}} \equiv -1 \mod N$, then fail
- otherwise return a factor $\gcd(N, a^{\frac{r}{2}}-1)$

We know from the number theory that the number of elements a that do not lead to failure is sufficient (at least one half). However, the impracticality of this algorithm stems from the fact that it is difficult to determine the order of the element in the group $\mathbb{Z}_N^*$. The quantum essence of Shor's algorithm is thus the search for the order of the element. For this task, the Fourier transform is suitable, and it is polynomial on a quantum computer.

Finding the order. The exponentiation of the element a modulo N, i.e. $k \mapsto a^k \mod N$, is the mapping $f : \mathbb{N} \to \mathbb{Z}_N^*$ with the period r. This gives a basic idea of why the Fourier transform can be useful for finding the order.

Quantum exponentiation must take place on finite binary registers. So let $n = \lceil \log N \rceil$ be the number of bits in the binary expansion of the number N, and choose some $M = 2^m$ large enough (the size of m will affect the probability of success of the algorithm).

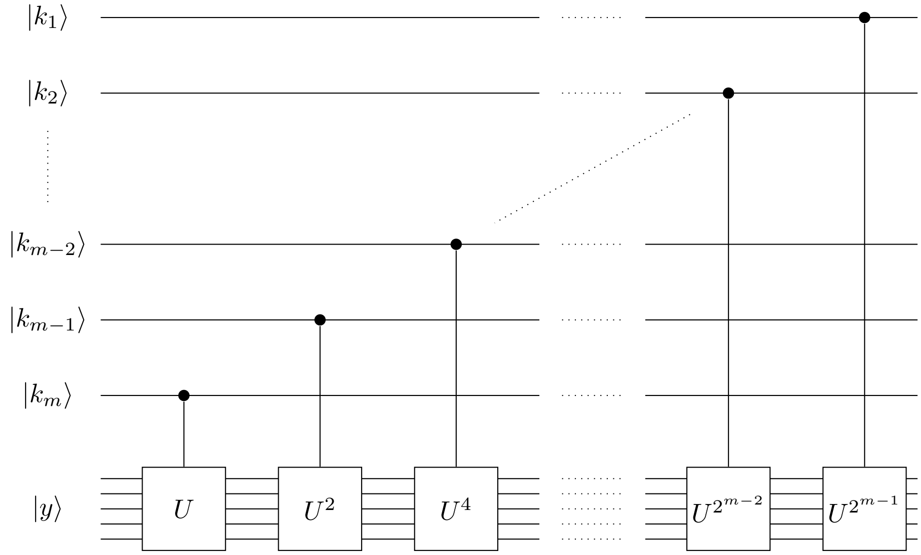
The exponentiation is simulated by the operator

$$\begin{aligned} W : \mathbb{H}_2^{\otimes m} \otimes \mathbb{H}_2^{\otimes n} &\to \mathbb{H}_2^{\otimes m} \otimes \mathbb{H}_2^{\otimes n} \\ |k\rangle|y\rangle &\mapsto |k\rangle|ya^k \mod N\rangle \end{aligned}$$

where for $N \leq y \leq 2^n - 1$, i.e. for elements for which the remainder would be repeated, we define $W|k\rangle|y\rangle := |k\rangle|y\rangle$. Because a relatively prime to N, the operator W permutes base elements and is therefore unitary. Implementation of the W operator is possible using modular exponentiation. If U is an operator for which we have controlled powers U^{2^j}, then the following circuit exponentiates U, that is, it realizes the mapping

$$|k\rangle|y\rangle \mapsto |k\rangle U^k|y\rangle,$$

in this way:

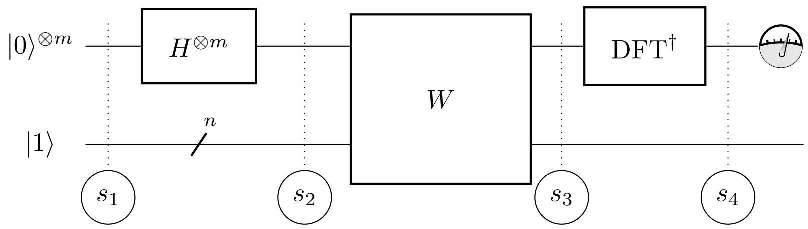


In the case of the operator W, U corresponds to the multiplication by the element a in the group $\mathbb{Z}_N$, i.e. the transformation

$$\begin{aligned} U : \mathbb{H}_2^{\otimes n} &\to \mathbb{H}_2^{\otimes n} \\ |y\rangle &\mapsto |ay \mod N\rangle, \end{aligned}$$

where again $U|y\rangle := |y\rangle$ pro $y \geq N$.

The basic idea of the order-revealing algorithm is the standard one: evaluate W on all values of $|k\rangle$ simultaneously. Because the exponentiation function is periodic, we apply the Fourier transform to it and we should get information about the period. The whole algorithm looks like this:



Note that the state $|1\rangle$ (or $|y\rangle$ for $y = 1$) is a base element of the n-qubit register with the number 1, ie $|0\rangle^{(n-1)}|1\rangle = |0\cdots01\rangle$. The first three phases give

$$s_1 : \ |0\rangle^{\otimes m}|0\cdots01\rangle \qquad s_2 : \ \frac{1}{\sqrt{M}}\sum_{k=0}^{M-1}|k\rangle|0\cdots01\rangle \qquad s_3 : \ \frac{1}{\sqrt{M}}\sum_{k=0}^{M-1}|k\rangle|a^k\rangle,$$

thereby preparing the desired uniform superposition of the values of the function $k \mapsto a^k$. By applying the Fourier transform (that is, IFT = DFT†) to the first register we get

$$s_4 : \quad \frac{1}{M}\sum_{k}^{M-1}\sum_{z}^{M-1}\exp\left[2\pi i\frac{kz}{M}\right]|z\rangle|a^k\rangle$$

We will now measure the first register. The probability that the measurement result will correspond to some selected $|z\rangle$ is obtained according to the postulate of the measurement as a square of the size of the projection on the subspace of the result, i.e. on the vector of components containing $|z\rangle$. It is thus the sum of the squares of the magnitude of the probability amplitudes for all terms in which $|z\rangle$ occurs. There are r such terms, namely $|z\rangle|a^0\rangle$, $|z\rangle|a^1\rangle$, …, $|z\rangle|a^{r-1}\rangle$, where the coefficient at $|z\rangle|a^t\rangle$ is the sum of the coefficients for all $|z\rangle|a^k\rangle$, where k is of the form $sr+t \mod N$. So all terms containing some fixed $|z\rangle$ are

$$\frac{1}{M}\sum_{t=0}^{r-1}\left(\sum_{s=0}^{\ell_t}\exp\left[2\pi i\frac{(sr+t)z}{M}\right]\right)|z\rangle|a^t\rangle$$

and the corresponding probability is

$$\begin{aligned} P(z) &= \frac{1}{M^2}\sum_{t=0}^{r-1}\left|\sum_{s=0}^{\ell_t}\exp\left[2\pi i\frac{(sr+t)z}{M}\right]\right|^2 = \\ &= \frac{1}{M^2}\sum_{t=0}^{r-1}\left|\exp\left[2\pi i\frac{tz}{M}\right]\right|^2\left|\sum_{s=0}^{\ell_t}\exp\left[2\pi i\frac{srz}{M}\right]\right|^2 = \\ &= \frac{1}{M^2}\sum_{t=0}^{r-1}\left|\sum_{s=0}^{\ell_t}\exp\left[2\pi i\frac{rz}{M}s\right]\right|^2 . \end{aligned}$$

The number ℓ_t is the largest such that $\ell_t r+t$ is less than M, i.e.

$$\ell_t = \left\lfloor\frac{M-1-t}{r}\right\rfloor .$$

The values of ℓ_t may differ by one for different t's. The complication is that r does not generally divide M; if it divided it, ℓ would simply be equal to $M/r-1$. This irregularity is of deeper importance. Note that we are performing a Fourier transform on the group $\mathbb{Z}_M$, not $\mathbb{Z}_N$! The result will have some inaccuracy, because the mapping $k \mapsto a^k \mod N$ is not completely periodic on $\mathbb{Z}_M$: around zero, the periodicity is broken (if r does not divide M). For large M, however, this inaccuracy will be negligible.

These general considerations are specified in the calculation of the value of $P(z)$. We will show that the following is true

$$(*)\qquad \left|\sum_{s=0}^{\ell_t}\exp\left[2\pi i\frac{rz}{M}s\right]\right| \approx \begin{cases} \frac{M}{r} & \text{if } rz\approx pM \text{ for some integer } p,\\ 0 & \text{otherwise.}\end{cases}$$

In the above-mentioned ideal case, where r divides M, the sum above ranges over values of some character of the group $\mathbb{Z}_M$, and the relation $(*)$ therefore holds with equality in the place of $\approx$. So we will measure z, which is of the form $p\cdot\frac{M}{r}$, where $p\in\{0,1,2,\dots,r-1\}$. For each such z, the probability $P(z)$ is equal to $\frac{1}{r}$, as is easily calculated. From z we obtain the fraction

$$\frac{z}{M}=\frac{p}{r},$$

whose denominator is r if p and r are coprime. This occurs for $r>19$ with a probability of at least $\frac{1}{4}\frac{1}{\log\log r}$. If p s r has a common factor, we get at least some factor of r. By repeating the procedure several times, we will most likely eventually obtain r.

In the general case, that is, if r does not divide M, the measured z is most likely close to some multiple of $\frac{M}{r}$, so that

$$\frac{z}{M} \approx \frac{p}{r}.$$

An interesting question arises as to how to find all fractions with a limited numerator that are close to a given value of α. The answer is the continued fraction expansion. It holds that if the distance between α and the fraction $\frac{p}{r}$ is less than $\frac{1}{2r^2}$, then this fraction is present in a continued fraction convergent of the number α (see the lecture in Czech on continued fractions within Number Theory and RSA, especially the application to Shor's algorithm on page 8, translated at the end of this chapter). If we assume that z is the rounded value of $p\frac{M}{r}$, that is, that

$$\left| z - p\frac{M}{r} \right| \leq \frac{1}{2},$$

then

$$\left| \frac{z}{M} - \frac{p}{r} \right| \leq \frac{1}{2M},$$

which leads to the choice of M to be approximately N^2 ensuring the detection of the corresponding $\frac{p}{r}$ using continued fractions.

It remains to show with what precision the estimate $(*)$ holds in these circumstances. Denote

$$\varphi = \frac{rz}{M} - p$$

the aproximation "error", which, according to our assumption, satisfies

$$|\varphi| \leq \frac{r}{2M}.$$

We approximate the sum of the geometric series (writing for simplicity ℓ instead of ℓ_t):

$$\left| \sum_{s=0}^{\ell} \exp\left[2\pi i \frac{rz}{M} s\right] \right|^2 = \left| \sum_{s=0}^{\ell} \exp\left[2\pi i \varphi s\right] \right|^2 = \frac{\left|\exp\left[2\pi i \varphi(\ell+1)\right] - 1\right|^2}{\left|\exp\left[2\pi i \varphi\right] - 1\right|^2} = \frac{\sin^2 \pi\varphi(\ell+1)}{\sin^2 \pi\varphi},$$

where the last equality follows from the relation

$$\left|e^{ix} - 1\right|^2 = (e^{ix} - 1)(e^{-ix} - 1) = 2(1 - \cos x) = 4\sin^2 \frac{x}{2}.$$

It is not difficult to verify that the value decreases with increasing φ, which is consistent with φ being a measure of inaccuracy: the maximum M/r is reached in our ideal case that corresponds to $\varphi = 0$. In addition, since $\sin^2$ is an even function, we get

$$\frac{\sin^2 \pi\varphi(\ell+1)}{\sin^2 \pi\varphi} \geq \frac{\sin^2 \frac{\pi}{2}\frac{r(\ell+1)}{M}}{\sin^2 \frac{\pi}{2}\frac{r}{M}}.$$

It follows from the definition of ℓ that $M - r < r(\ell+1) < M + r$. The numerator of the fraction is therefore very close to one (for $r/M < 1/100$ differs from one by less than a thousandth) and the denominator, which, on the other hand, is very small, can be upper bounded quite accurately by the relation $\sin x < x$. In total we get

$$\left|\sum_{s=0}^{\ell} \exp\left[2\pi i \frac{rz}{M} s\right]\right|^2 > 0.999 \cdot \frac{4}{\pi^2}\frac{M^2}{r^2} > \frac{2}{5}\frac{M^2}{r^2}$$

and

$$P(z) > \frac{2}{5}\frac{1}{r}.$$

We can conclude that with a probability of at least $\frac{2}{5}$ we measure z, for which is $\frac{p}{r}$ present in the continued fraction expansion of $\frac{z}{M}$.

The overall success rate of the algorithm is summarized in the following table:

success condition	probability
choosing a suitable a	$\frac{1}{2}$
z is close to $p\frac{M}{r}$	$\frac{2}{5}$
p is coprime with r	$\frac{1}{4}\frac{1}{\log\log n}$

So the total success rate is at least $\frac{1}{20}\frac{1}{\log\log n}$. E.g. for the RSA module of length 4096, the success rate of one round of the algorithm is at least 0.6%, so four hundred rounds gives more than 90% probability of success. This estimate is unnecessarily pessimistic especially in the requirement that r and $p$are coprime; even if r and p have common factors, we get some of them in each round and after several attempts it is likely to reconstruct r as the least common multiple of the factors found.

*

Example from the lecture on RSA. Continued fractions are an effective tool for the rational approximation of irrational numbers. However, they are also important for the approximation of rational numbers. Suppose we have an inaccurate value of a fraction, caused by, for example, rounding or measurement inaccuracy. An example of such a situation is Shor's quantum factorization algorithm. To reveal the original fraction, we use the continued fraction expansion of an inaccurate value.

Example: We have the value $h = 0.15328$, which we know is the rounding (to the nearest hundredth of a thousand) of a proportion of at most eight-bit numbers. The continued fraction expansion of h is $[0, 6, 1, 1, 9, 1, 10]$ with convergents:

$$\left(0, \frac{1}{6}, \frac{1}{7}, \frac{2}{13}, \frac{19}{124}, \frac{21}{137}, \frac{229}{1494}, \frac{479}{3125}\right).$$

Of the fractions with a denominator and a numerator of at most eight bits, only $^{21}/_{137}$ is equal to h when rounded to the nearest hundredth of thousand.

zlomek	zaokrouhlení
$\frac{1}{6}$	0.16667
$\frac{1}{7}$	0.14286
$\frac{2}{13}$	0.15385
$\frac{19}{124}$	0.15323
$\frac{21}{137}$	0.15328

Of course, the question arises as to whether we have not missed a fraction with the same rounding in the continued fraction. The following statement is relevant to this question.

Theorem: If

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

then the fraction $^{p}/_{q}$ is a convergent of α.

In the above example, the denominator is less than 256 and the rounding error is at most $5 \cdot 10^{-6}$. Because

$$5 \cdot 10^{-6} < \frac{1}{2 \cdot 256^2},$$

we see that the fraction sought is indeed one of the convergents.