The main piece is the Quantum Fourier Transform, which is just the Discrete Fourier Transform with exponential scaling:

5.2 Fourier Matrices

The next important family consists of the quantum Fourier matrices. Let ω stand for $e^{2\pi i/N}$, which is often called "the" principal Nth root of unity.

DEFINITION 5.2 The Fourier matrix $\boldsymbol{F}_N$ of order N is

$$\frac{1}{\sqrt{N}}\begin{bmatrix} 1 & 1 & 1 & 1 & \cdots & 1 \\ 1 & \omega & \omega^2 & \omega^3 & \cdots & \omega^{N-1} \\ 1 & \omega^2 & \omega^4 & \omega^6 & \cdots & \omega^{N-2} \\ 1 & \omega^3 & \omega^6 & \omega^9 & \cdots & \omega^{N-3} \\ \vdots & & & \vdots & \ddots & \vdots \\ 1 & \omega^{N-1} & \omega^{N-2} & \omega^{N-3} & \cdots & \omega \end{bmatrix}$$

That is, $\boldsymbol{F}_N[i,j] = \omega^{ij \bmod N}$ divided by $\sqrt{N}$.

It is well known that $\boldsymbol{F}_N$ is a unitary matrix over the complex Hilbert space. This and further facts about $\boldsymbol{F}_N$ are set as exercises at the end of this chapter, including a running theme about its feasibility via various decompositions. For any vector $\boldsymbol{a}$, the vector $\boldsymbol{b} = \boldsymbol{F}_N\boldsymbol{a}$ is defined in our index notation by

$$\boldsymbol{b}(x) = \frac{1}{\sqrt{N}} \sum_{t=0}^{N-1} \omega^{xt} \boldsymbol{a}(t).$$

Compare-contrast with the Hadamard Transform:

H	0000	0001	0010	0011	0100	0101	0110	0111	1000	1001	1010	1011	1100	1101	1110	1111
0000	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
0001	1	−1	1	−1	1	−1	1	−1	1	−1	1	−1	1	−1	1	−1
0010	1	1	−1	−1	1	1	−1	−1	1	1	−1	−1	1	1	−1	−1
0011	1	−1	−1	1	1	−1	1	−1	1	−1	−1	1	1	−1	1	−1
0100	1	1	1	1	−1	−1	−1	−1	1	1	1	1	−1	−1	−1	−1
0101	1	−1	1	−1	−1	1	−1	1	1	−1	1	−1	−1	1	−1	1
0110	1	1	−1	−1	−1	−1	1	1	1	1	−1	−1	−1	−1	1	1
0111	1	−1	−1	1	−1	1	1	−1	1	−1	−1	1	−1	1	1	−1
1000	1	1	1	1	1	1	1	1	−1	−1	−1	−1	−1	−1	−1	−1
1001	1	−1	1	−1	1	−1	1	−1	−1	1	−1	1	−1	1	−1	1
1010	1	1	−1	−1	1	1	−1	−1	−1	−1	1	1	−1	−1	1	1
1011	1	−1	−1	1	1	−1	1	−1	−1	1	1	−1	−1	1	1	−1
1100	1	1	1	1	−1	−1	−1	−1	−1	−1	−1	−1	1	1	1	1
1101	1	−1	1	−1	−1	1	−1	1	−1	1	−1	1	1	−1	1	−1
1110	1	1	−1	−1	−1	−1	1	1	−1	−1	1	1	1	1	−1	−1
1111	1	−1	−1	1	−1	1	1	−1	−1	1	1	−1	1	−1	−1	1

$$\mathbf{H}[u,v] \;=\; (-1)^{u \bullet v}$$

The centerpiece of Peter Shor's algorithm detects a *period* in a function. Let

$$f\colon \mathbb{N} \to \{\,0, 1, \ldots, M-1\,\}$$

be a feasibly computable function. We are promised that there is a period r, meaning that, for all x,

$$f(x+r) = f(x).$$

The goal is to detect the period, that is, to determine the value of r. Actually, we need more than this promise. We also need that the repeating values

$$f(0), f(1), \ldots, f(r-1)$$

are all distinct. Some call this latter condition "injectivity" or "bijectivity." Possible relaxations of this condition are explored in the exercises, and overall its necessity and purpose are not fully understood.

1. Given an n-bit integer M, which we suppose is a product of distinct primes, use classical randomness to generate an integer a between 1 and $M-1$. First, we check for the tiny chance that a already shares a factor with M, in which case one application of Euclid's algorithm for the greatest common divisor quickly finds it and we're done. Otherwise, form the function $f_a(x) = a^x \bmod M$, which then has a period r that we wish to compute. This r will divide the product of $p-1$ over all primes p dividing M and *may* help

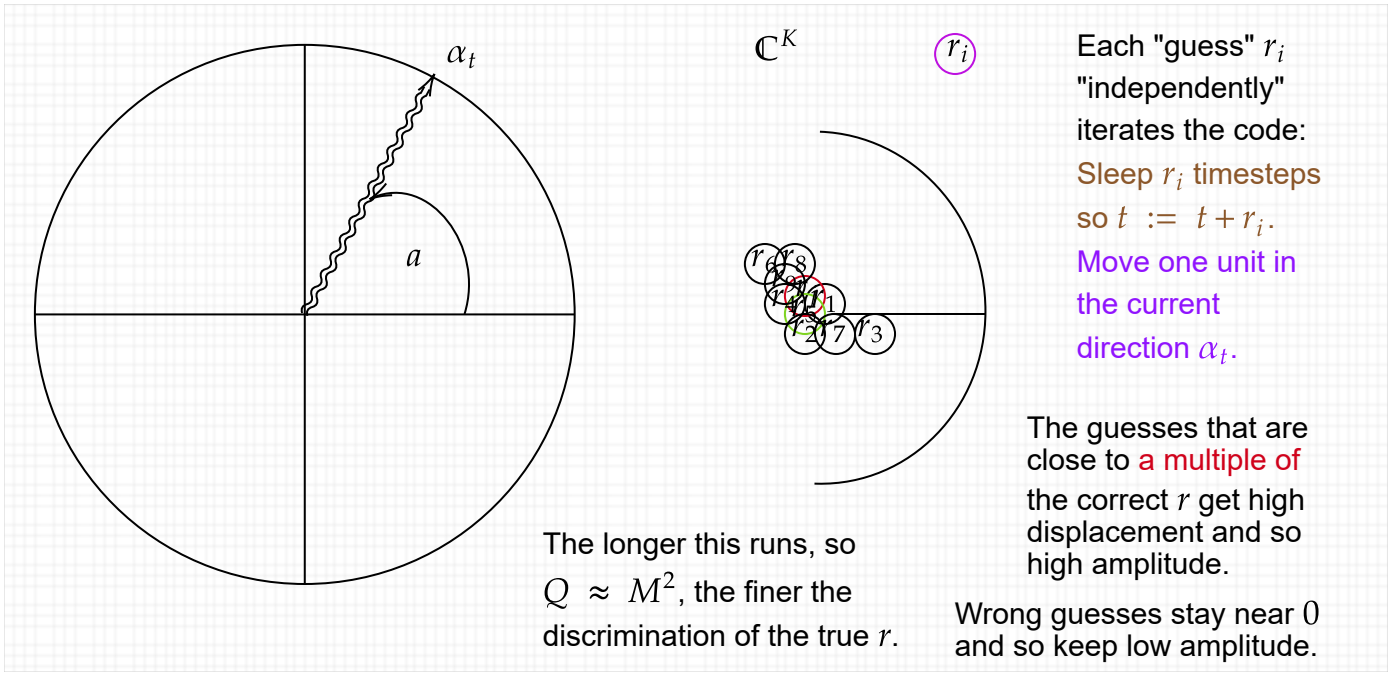
us find them. Many values of a and r are unhelpful, but with substantial probability, a will be chosen so that r is computed and yields a factor p.

2. Use the classical feasibility of modular exponentiation via repeated squaring (shown in problems 2.8–2.9) to prepare the functional superposition of $f_a(x)$ over all $x < M$.
3. Run the quantum part once and measure all qubits. The string formed by the first ℓ of them, where ℓ is about $2 \log_2 M$, yields a particular integer x in binary encoding. With substantial probability, x is "good," as defined below.
4. Then classical computation is used to try to infer r from x. Either this succeeds and we go to the next step, or it is recognized that x is not good. In the latter case, we go back to step 3, running the quantum routine again.
5. There is still a chance the value of r may be unsuitable—that is, that the original a was an unlucky choice. In this case, we must begin again at step 1. But otherwise, the value of r provides the only needed input to a final classical stage that yields a verifiable solution to the problem about M.

$$\Phi_f = \frac{1}{\sqrt{N}} \sum_{x \in \{0,1\}^q} |x f(x)\rangle$$

The Intuition (See also Scott Aaronson, https://www.scottaaronson.com/blog/?p=208)

Let r stand for the true period of f. Let a be any element of the group G of size $\phi(M) = (p-1)(q-1)$. Then we will picture a as a "crazy clock" that jumps a units *counter*-clockwise at each time step.



With fairly high probability, measurement yields a multiple of r. The true r is the least of the multiples. It is individually the most likely value returned and is also returned with reasonable probability. A bad r might work anwyay. We can tell whether r works by seeing if the classical part gives us p or q, else we just try the quantum process again.

Arora-Barak uses cubic slack, $Q = 2^m$ where $m \doteq 3 \log M$.

Details

The top-down goal is to find a number X such that $X^2 \equiv 1$ modulo M but X is not $\equiv 1$ or $\equiv -1$ modulo M. Then $X^2 - 1 = (X-1)(X+1)$ is a multiple of M but neither factor is zero. When $M = pq$ with p, q prime, this means p and q each divide one or both factors. We need to split them across the factors, so that $\gcd(X-1, M)$ and/or $\gcd(X+1, M)$ will find p and q as opposed to just giving M back again. Thus we want to guess a such that:

1. The period r of a is even, so that $r/2$ is defined;
2. $X = a^{r/2} \not\equiv M - 1$ modulo M.
3. Either $X - 1$ or $X + 1$ is a multiple of one of p, q **but not both**.

If our value of a fails either of these, we just try again from the start of guessing $a < M$.

Our treatment (blog post and chapter 12) also desires r to be a multiple of $p-1$ or $q-1$. It can be shown that many a give this "helpful" property, which requires $r \geq \sqrt{(p-1)(q-1)} \approx \sqrt{M}$.

(It is not clear whether we show this. It could be an exercise: Consider numbers r that divide a product mn of two nearly-equal composite numbers. Conditioned on $r \geq \min\{m, n\}$, give a lower bound for the proportion that are a multiple of m or a multiple of n. Note that m and n need not be themselves relatively prime; $p-1$ and $q-1$ are both even, for instance. It would still need to be argued that most a give such an r. But I am not sure that the "helpful" property is needed either.)

Chapter 12 does handle the argument in property 3, given that r is "helpful"---which also subsumes issue 1 since $p-1$ and $q-1$ are even. Issue 2 is handled by a random argument.

We will see that the closer r is to $\sqrt{M}$ as opposed to being order-of M, the more challenging for a potential classical simulation of Shor's algorithm.

Another thing to observe is that when M is a Blum integer, meaning p and q are both congruent to 3 modulo 4, then $(p-1)(q-1)$ is divisible by 4 but no higher even number. There are always four square roots of 1 modulo $M = pq$, so we need to argue that the a's such that $a^{r/2}$ is one of the good ones are as plentiful as the bad ones. (Note that r depends only on a.) Here is an example for the smallest Blum integer: $21 = 3*7$. The quadratic residues are:

$1:1, 2:4, 3:9, 4:16, 5:4, 6:15, 7:7, 8:1, 9:18, 10:16,$
$20:1, 19:4, 18:9, 17:16, 15:15, 14:7, 13:1, 12:18, 11:16$

Now $(p-1)(q-1) = 12$. The numbers $Y = 8-1, 8+1, 13+1$, and $13-1$ all give a factor via $\gcd(21, Y)$.

$a = 1$: $r = 1$; of course doesn't work.
$a = 2$: $2, 4, 8, 16, 11, 1$. Works
$a = 4$: $16, 1$ (period 3 is odd)
$a = 5$: $4, 20, 16, 17, 1$; doesn't work because $20 \equiv -1$.
$a = 8$: $8^2 \equiv 1$. Period $r = 2$ is "helpful" and $8^{r/2} = 8$ is not -1. So works.
$a = 10$: $16, 13, 4, 19, 1$. Works
The other values are mirror images.

A more interesting Blum integer IMHO is $77 = 7*11$. Then $(p-1)(q-1) = 60$. "Helpful" means the period is a multiple of 6 or of 10. Note: $34^2 = 1156 = 77*15+1$ is a nontrivial square root of 1 and $43^2 = 1849 = 77*24+1$ is the other one. Does 2 work?

$2:4, 8, 16, 32, 64, 51, 25, 50, 23, 46, 15, 30, 60, 43, 9, 18, 36, 72, 67, 57, 37, 74,$ etc.: yes.

The next question is whether it is OK for the quantum part to obtain a multiple $r' = br$ of a helpful r. If b is even than certainly not, because $a^{r'/2}$ will be 1. But if b is odd---? In any event, we can obviate this question because we can single out the minimum r with sufficiently high probability.

The key auxiliary technical notion is a number x that is "good" to help find r.

11.2 Good Numbers

Let Q be a power of two, $Q = 2^\ell$, such that $M^2 \le Q < 2M^2$. Say an integer x in the range $0, 1, \ldots, Q-1$ is **good** provided there is an integer t relatively prime to the period r such that

$$tQ - xr = k, \quad \text{where} \quad -r/2 \le k \le r/2. \tag{11.1}$$

The key part is the multiple t of Q being relatively prime to r.

LEMMA 11.1 There are $\Omega\left(\frac{r}{\log\log r}\right)$ good numbers.

Proof. The key insight is to think of equation (11.1) as an equation modulo r. Then it becomes

$$tQ \equiv k \bmod r,$$

where $-r/2 \le k \le r/2$. But as t varies from 0 to $r-1$, the value of k can be arranged to be always in this range, so the only constraint on t is that it must be relatively prime to r. The number of values t that are relatively prime to r defines Euler's *totient* function, which is denoted by $\phi(r)$. Note that for each value of t there is a different value of x, so counting ts is the same as counting xs. Thus, the lemma reduces to a lower bound on Euler's function. But it is known that

$$\phi(z) = \Omega\left(\frac{z}{\log\log z}\right).$$

Indeed, the constant in Ω approaches $e^{-\gamma}$, where $\gamma = 0.5772156649\ldots$ is the famous Euler-Mascheroni constant. In any event, this proves the lemma. □

The general drift is that a good x gives a good chance of finding r exactly, by purely classical means. Of note:

If r is close to M, then by choosing Q close to M rather than M^2, we would stand a good chance of finding a good x just by picking about $\log \ell$-many of them classically at random. However, this does not help when r is smaller. The genius of Shor's algorithm is that the quantum Fourier transform can be used to drive amplitude toward good numbers in all cases.

This makes $r \approx M^{1-\epsilon}$ where $0 < \epsilon < 1$ the "vat" of hard cases: too sparse to guess at random. For the quantum part, however, we need $Q > rM$. Just to finish off the classical part:

LEMMA 11.7 If x is good, then in classical polynomial time, we can determine the value of r.

Proof. Recall that x being good means that there is a t relatively prime to r so that (by symmetry)

$$xr - tQ = k \qquad \text{where} \qquad -\frac{r}{2} \le k \le \frac{r}{2}.$$

Assume that $k \ge 0$; the argument is the same in the case where it is negative. We can divide by rQ and get the equation

$$\left| \frac{x}{Q} - \frac{t}{r} \right| \le \frac{1}{2Q}.$$

We next claim that r and t are unique. Suppose there is another t'/r'. Then

$$\left| \frac{t}{r} - \frac{t'}{r'} \right| \ge \frac{1}{rr'} \ge \frac{1}{M^2}.$$

But then both fractions are close, which makes Q smaller than M^2, a contradiction.

Because r is unique, it follows that t is too. So we can treat

$$xr - tQ = k$$

as an integer program in a fixed number of variables: the variables are r, t, and two slack variables used to state

$$-r/2 \le k \le r/2$$

as two equations. While integer programs are hard in general, for a fixed number of variables they are solvable in polynomial time. This proves the lemma. □

The Quantum Part

The essence is to show that a non-negligible amount of amplitude is marshaled onto "good" values x. We actually analyze pairs xy formed by the main and ancilla qubits after a one-shot sampling measurement.

1. The start vector $\boldsymbol{a}$ is the functional superposition of f; that is,

$$\boldsymbol{a}(xy) = \begin{cases} \frac{1}{\sqrt{Q}} & \text{when } y = f(x); \\ 0 & \text{otherwise.} \end{cases}$$

2. The next vector $\boldsymbol{b}$ is the result of applying $\boldsymbol{F}_Q$ to the x part of $\boldsymbol{a}$.
3. Measure $\boldsymbol{b}$, giving an answer xy from which we discard y.
4. Exit into a classical routine that tests whether x is a good integer—if so, continue with the classical steps given later or else repeat from step 1.

Recall that xr in equation (11.1) is ordinary numerical multiplication, whereas xy in vector indices is binary string concatenation. Although y is discarded, the injectivity condition ensures that for every good x the superposition caused by $\boldsymbol{F}_Q$ will contribute exactly r-many ys. Together with lemma 11.1, this will give a little short of order-r^2 good *pairs* xy. Hence, it suffices to show that every good pair receives $\Omega\left(\frac{1}{r}\right)$ of the amplitude, giving $\Omega\left(\frac{1}{r^2}\right)$ in probability. The resulting $\Omega\left(\frac{1}{\log \log r}\right)$ probability of getting a good number on each trial will be large enough for the classical part to expect to succeed after relatively few trials of the quantum part.

The intuition is that the quantum Fourier transform creates power series out of many angles β. Each series creates a large locus of points $0, \beta, 2\beta, 3\beta, \ldots$ For most angles β, the locus spreads itself over the circle so that its average—which is obtained by summing the corresponding power series of complex numbers $\exp(ik\beta)$—is close to the origin. If β is close to an integer multiple of 2π radians, however, then the angles all stay close to 0 modulo 2π, and the average stays close to the complex number 1. These "good" β embody multiples of the unknown period r, so the process will distinguish those x that yield such β. The way that r "pans out" like a nugget of gold is similar to what happens with s in Simon's algorithm.

Much of the analysis can be done exactly before we take estimates to bound the amplitudes. It suffices to show that the good cases collectively grab a non-trivial fraction of the probability—we do not need estimates when β is "bad" at all. Let us consider any pair xy where y is in the range of f. With $\omega = \exp\left(\frac{2\pi i}{Q}\right)$, we have

$$\boldsymbol{b}(xy) = \frac{1}{\sqrt{Q}} \sum_{u=0}^{Q-1} \omega^{xu} \boldsymbol{a}(uy) = \frac{1}{\sqrt{Q}} \sum_{u:f(u)=y} \omega^{xu} \boldsymbol{a}(uy) = \frac{1}{Q} \sum_{u \in f^{-1}(y)} \omega^{xu}.$$

The last $\frac{1}{Q}$ is not a typo—we have substituted the value of $\boldsymbol{a}(uy)$. Now take the first x_0 such that $f(x_0) = y$. Then, by injectivity,

$$f^{-1}(y) = \{x_0,\ x_0 + r,\ x_0 + 2r,\ x_0 + 3r, \ldots\}.$$

The cardinality of this set (up to $Q-1$) is $T = 1 + \lfloor \frac{Q-x_0}{r} \rfloor$. This brings out the finite geometric series and enables us to apply the formula for its sum:

$$\boldsymbol{b}(xy) = \frac{1}{Q} \sum_{k=0}^{T-1} \omega^{x(x_0+rk)} = \frac{\omega^{xx_0}}{Q} \sum_{k=0}^{T-1} \omega^{xrk} = \omega^{xx_0} \frac{1}{Q} \left(\frac{\omega^{Txr} - 1}{\omega^{xr} - 1} \right).$$

Note that when we take absolute values, the complex-phase factor ω^{xx_0} will go away because it is a unit. We can multiply by further such units to make the numerator and denominator have real values even before we take the norms, using the trick that $\exp(i\beta) - \exp(-i\beta) = 2\sin(\beta)$:

$$\begin{aligned}
\boldsymbol{b}(xy) &= \omega^{xx_0 - xrT/2} \frac{1}{Q} \left(\frac{\omega^{Txr/2} - \omega^{-Txr/2}}{\omega^{xr} - 1} \right) \\
&= \omega^{xx_0 - xrT/2 + xr/2} \frac{1}{Q} \left(\frac{\omega^{Txr/2} - \omega^{-Txr/2}}{\omega^{xr/2} - \omega^{-xr/2}} \right) \\
&= \omega^{x(x_0 + (T-1)r)} \frac{1}{Q} \left(\frac{\sin(T \cdot \pi xr/Q)}{\sin(\pi xr/Q)} \right).
\end{aligned}$$

Note that we canceled a factor of 2 both inside and outside the angles. This finally tells us that

$$|\boldsymbol{b}(xy)|^2 = \frac{1}{Q^2} \frac{\sin^2(T \cdot \pi xr/Q)}{\sin^2(\pi xr/Q)}. \tag{11.2}$$

looks strange that the right-hand side is independent of y, but recall that we did use the property that y is in the range of f—without needing that $y=f(x)$. By injectivity, we have r-many such ys for any particular x. To finish the analysis, we need to show the following:

1. When x is good, the right-hand side of equation (11.2) is relatively large.
2. The total probability on good *pairs* xy is $\Omega\left(\frac{1}{\log n}\right)$, where n is the number of digits in M, which is high enough to give high probability of finding a good x in $O(\log n)$-many trials.
3. If x is good, then in classical polynomial time we can determine the value of r.

The second statement will follow quickly after the first, and we handle both in the next section.

11.5 Probability of a Good Number

We state a fact about sines that has its own interest. Note that 1.581 in radians is a little bit more than $\pi/2$ to leave some slack. We target the number 0.63247 because its square is just above 0.4.

LEMMA 11.2 For all $T > 0$ and all angles $\alpha > 0$ such that $T\alpha \leq 1.581$,

$$\frac{\sin(T\alpha)}{\sin(\alpha)} > 0.63247T.$$

Proof. The well-known identity $\sin(\alpha) \leq \alpha$, which holds for all $\alpha \geq 0$, makes it suffice to show that

$$\frac{\sin(T\alpha)}{T\alpha} > 0.63247.$$

Consider the function $\frac{\sin(x)}{x}$ for $0 < x \leq 1.581$. Its derivative has numerator $x\cos(x) - \sin(x)$ and denominator x^2. For $\frac{\pi}{2} < x < 1.581$ the derivative is negative since $\cos(x)$ is negative. For $0 < x < \frac{\pi}{2}$ it is also negative because the inequality $x < \tan(x)$ holds there. Because its derivative is always negative in this range, the function is minimized at the upper boundary $x = 1.581$, where it has value

$$\frac{\sin(1.581)}{1.581} > \frac{0.9999479}{1.581} > 0.63247.$$

□

LEMMA 11.3 For all pairs xy with x good, assuming $M \geq 154$, the probability of the measurement step outputting x is bounded below by $\frac{0.4}{r^2}$.

Proof. Recall that x being good means there is an integer t such that $-\frac{r}{2} \leq tQ - xr \leq \frac{r}{2}$, and that we have $Q > Mr$ and $T = 1 + \lfloor \frac{Q - x_0}{r} \rfloor$, where $x_0 \leq r$. From above, using $|\sin(x)| = |\sin(-x)| = |\sin(x + \pi)|$, we have

$$\begin{aligned}
|\boldsymbol{b}(xy)|^2 &= \frac{1}{Q^2} \frac{\sin^2(T \cdot \pi xr/Q)}{\sin^2(\pi xr/Q)} \\
&= \frac{1}{Q^2} \frac{\sin^2\left(T \cdot \pi(\frac{xr}{Q} - t)\right)}{\sin^2\left(\pi(\frac{xr}{Q} - t)\right)} \\
&= \frac{1}{Q^2} \frac{\sin^2\left(T \cdot \pi \frac{xr - tQ}{Q}\right)}{\sin^2\left(\pi \frac{xr - tQ}{Q}\right)} \\
&= \frac{1}{Q^2} \frac{\sin^2\left(T \cdot \pi \frac{tQ - xr}{Q}\right)}{\sin^2\left(\pi \frac{tQ - xr}{Q}\right)} .
\end{aligned}$$

Now, by goodness, the angle $\alpha = \pi \frac{tQ - xr}{Q}$ is at most $\pi \frac{r}{2Q}$. Because $T \leq 1 + \frac{Q}{r}$, we have

$$T\alpha \leq \frac{\pi}{2} + \frac{\pi r}{2Q}.$$

Because we chose $Q > Mr$, we have $\frac{\pi r}{2Q} < \frac{\pi}{2M} < 1.581 - \frac{\pi}{2}$ using the condition $M \geq 154$. Thus, $T\alpha \leq 1.581$, thus meeting the hypothesis of lemma 11.2. This gives us what we needed to hit our round-number probability target:

$$|\boldsymbol{b}(xy)|^2 \geq \frac{1}{Q^2}(0.63247T)^2 > \frac{0.4}{r^2}.$$

□

COROLLARY 11.4 The probability of getting a good number on each trial of the quantum part is $\Omega\left(\frac{1}{\log \log M}\right)$—indeed, at least 1 in $\log_2 \log_2 M$.

Proof. For every good x, there are r-many different ys for which $f^{-1}(y)$ is a set of cardinality T in the analysis of section 11.4. Thus, by lemma 11.1, there are $\Omega\left(\frac{r^2}{\log \log r}\right)$ *pairs* xy for which lemma 11.3 applies. A glance at the proof plus converting natural logs to logs base 2 makes the number of such pairs at least $\frac{2.53r^2}{\log_2 \log_2 r}$. Thus, the total probability of getting a good number is at least

$$\frac{2.53}{2.5 \log_2 \log_2 r} > \frac{1}{\log_2 \log_2 M}.$$

□

This finishes the analysis of (the quantum part of) Shor's algorithm. ☒