Scalability of Shor's algorithm with a limited set of rotation gates

Austin G. Fowler and Lloyd C. L. Hollenberg
Centre for Quantum Computer Technology,
School of Physics, University of Melbourne,
Victoria 3010, AUSTRALIA.
(Dated: April 1, 2022)

Typical circuit implementations of Shor’s algorithm involve controlled rotation gates of magnitude \( \pi/2^L \) where \( L \) is the binary length of the integer \( N \) to be factored. Such gates cannot be implemented exactly using existing fault-tolerant techniques. Approximating a given controlled \( \pi/2^\ell \) rotation gate to within \( \delta = O(1/2^\ell) \) currently requires both a number of qubits and number of fault-tolerant gates that grows polynomially with \( d \). In this paper we show that this additional growth in space and time complexity would severely limit the applicability of Shor’s algorithm to large integers. Consequently, we study in detail the effect of using only controlled rotation gates with \( d \) less than or equal to some \( d_{\text{max}} \). It is found that integers up to length \( L_{\text{max}} = O(4^{d_{\text{max}}}) \) can be factored without significant performance penalty implying that the cumbersome techniques of fault-tolerant computation only need to be used to create controlled rotation gates of magnitude \( \pi/64 \) if integers thousands of bits long are desired factored. Explicit fault-tolerant constructions of such gates are also discussed.

I. SHOR’S ALGORITHM

Shor’s algorithm factors an integer \( N = N_1N_2 \) by finding the period \( r \) of a function \( f(k) = m^k \mod N \) where \( 1 < m < N \) and \( \gcd(m, N) = 1 \). Provided \( r \) is even and \( f(r/2) \neq N - 1 \) the factors are \( N_1 = \gcd(f(r/2) + 1, N) \) and \( N_2 = \gcd(f(r/2) - 1, N) \), where \( \gcd \) denotes the greatest common divisor. The probability of finding a suitable \( r \) given a randomly selected \( m \) such that \( \gcd(m, N) = 1 \) is greater than 0.75 \[1\]. Thus on average very few values of \( m \) need to be tested to factor \( N \).

The quantum heart of Shor’s algorithm can be viewed as a subroutine that generates numbers of the form \( j = \pm 2^{2L}/r \). To distinguish this from the necessary classical pre- and postprocessing, this subroutine will be referred to as QPF (quantum period finding). For physical reasons, the probability \( s \) that QPF will successfully generate useful data may be quite low with many repetitions required to work out the period \( r \) of a given \( f(k) = m^k \mod N \). Using this terminology, Shor’s algorithm consists of classical preprocessing, potentially many repetitions of QPF with classical postprocessing and possibly a small number of repetitions of this entire cycle. This cycle is summarized in Fig.1.

A number of different quantum circuits implementing QPF have been designed \[3\, \[4\, \[5\, \[6\, \[7\, \[8\]. Table 1 summarizes the number of qubits required and the depth of each of these circuits. The depth of a circuit has been defined to be the minimum number of 2-qubit gates that must be applied sequentially to complete the circuit. It has been assumed that multiple disjoint 2-qubit gates can be implemented in parallel, hence the total number of 2-qubit gates can be significantly greater that the depth. For example, the Beaufregard circuit has a 2-qubit gate count of \( 8L^4 \) to first order in \( L \). Note that in general the depth of the circuit can be reduced at the cost of additional qubits.

The underlying algorithm common to each circuit begins by initializing the quantum computer to a single
Select $1 < m < N$ such that $\gcd(m, N) = 1$ (classical)

Try to find $j = c^{2L/r}$ (quantum)

Try to use $j$ to find period $r$ of $f(k) = m^k \mod N$ (classical)

Test whether $r$ is even and $m^{r/2} \mod N \not\equiv \pm 1 \mod N$ (classical)

$N_1 = \gcd(m^{r/2} - 1, N)$
$N_2 = \gcd(m^{r/2} + 1, N)$

Success

Fail

FIG. 1: The complete Shor’s algorithm including classical pre- and postprocessing. The first branch is highly likely to fail, resulting in many repetitions of the quantum heart of the algorithm, whereas the second branch is highly likely to succeed.

| Circuit  | Qubits | Depth |
|----------|--------|-------|
| Beauregard | $7$ | $2L$ | $32L^2$ |
| Vedral    | $5$ | $5L$ | $240L^3$ |
| Zalka     | $5$ | $\approx 50L$ | $2^{17}L^2$ |
| Gossett   | $4$ | $O(L^2)$ | $O(L \log L)$ |

Note that this step requires additional ancilla qubits. The exact number depends heavily on the circuit used.

Step four can actually be omitted but it explicitly shows the origin of the period $r$ being sought. Measuring the $f$-register yields

$$\frac{\sqrt{r}}{2L} \sum_{n=0}^{2^{2L}/r-1} |k_0 + nr\rangle_{2L} |f_M\rangle_L$$

where $k_0$ is the smallest value of $k$ such that $f(k)$ equals the measured value $f_M$.

Step five is to apply the quantum Fourier transform

$$|k\rangle \rightarrow \frac{1}{2^L} \sum_{j=0}^{2^{2L}-1} \exp\left(\frac{2\pi i}{2^L} jk\right) |j\rangle$$

to the $k$-register resulting in

$$\frac{\sqrt{r}}{2^L} \sum_{j=0}^{2^{2L}-1} \sum_{p=0}^{2^L/r-1} \exp\left(\frac{2\pi i}{2^L} (jk_0 + jr)\right) |j\rangle_{2L} |f_M\rangle_L.$$ (5)

The probability of measuring a given value of $j$ is thus

$$\Pr(j, r, L) = \left| \frac{\sqrt{r}}{2^L} \sum_{p=0}^{2^{2L}/r-1} \exp\left(\frac{2\pi i}{2^L} jpr\right) \right|^2.$$ (6)

If $r$ divides $2^{2L}$ Eq (6) can be evaluated exactly. In this case the probability of observing $j = c^{22L}/r$ for some integer $0 \leq c < r$ is $1/r$ whereas if $j \neq c^{22L}/r$ the probability is $0$. This situation is illustrated in Fig 2(a). However if $r$ divides $2^{2L}$ exactly a quantum computer is not needed as $r$ would then be a power of 2 and easily calculable. When $r$ is not a power of 2 the perfect peaks of Fig 2(a) become slightly broader as shown in Fig 2(b). All one can then say is that with high probability the value $j$ measured will satisfy $j \approx c^{2L}/r$ for some $0 \leq c < r$.

Given a measurement $j \approx c^{2L}/r$ with $c \neq 0$, classical postprocessing is required to extract information about $r$. The process begins with a continued fraction expansion. To illustrate, consider factoring 143 ($L = 8$). Suppose we choose $m$ equal 2 and the output $j$ of QPF is 31674. The relation $j \approx c^{2L}/r$ becomes $31674 \approx c65536/r$. The continued fraction expansion of $c/r$ is

$$\frac{31674}{65536} = \frac{1}{2 + \frac{1}{2 + \frac{1}{2 + \frac{1}{2 + \frac{1}{2 + \frac{1}{10 + \frac{1}{52}}}}}}}.$$ (7)

The continued fraction expansion of any number between 0 and 1 is completely specified by the list of denominators which in this case is $\{2, 14, 2, 10, 52\}$. The $n$th convergent of a continued fraction expansion is the proper fraction equivalent to the first $n$ elements of this list. An introductory exposition and further properties of continued
fractions are described in Ref. 4.

\[
\{2\} = \frac{1}{2}
\]
\[
\{2, 14\} = \frac{14}{29}
\]
\[
\{2, 14, 2\} = \frac{29}{60}
\]
\[
\{2, 14, 2, 10\} = \frac{304}{629}
\]
\[
\{2, 14, 2, 10, 52\} = \frac{15837}{32768}
\]

The period \( r \) can be sought by substituting each denominator into the function \( f(k) = 2^k \mod 143 \). With high probability only the largest denominator less than \( 2^L \) will be of interest. In this case \( 2^{60} \mod 143 = 1 \) and hence \( r = 60 \).

Two modifications to the above are required. Firstly, if \( c \) and \( r \) have common factors, none of the denominators will be the period but rather one will be a divisor of \( r \). After repeating QPF a number of times, let \( \{ j_m \} \) denote the set of measured values. Let \( \{ c_{mn}/d_{mn} \} \) denote the set of convergents associated with each measured value \( \{ j_m \} \). If a pair \( c_{mn}, c_{m'n'} \) exists such that \( \text{gcd}(c_{mn}, c_{m'n'}) = 1 \) and \( d_{mn}, d_{m'n'} \) are divisors of \( r \) then \( r = \text{lcm}(d_{mn}, d_{m'n'}) \), where lcm denotes the least common multiple. It can be shown that given any two divisors \( d_{mn}, d_{m'n'} \) with corresponding \( c_{mn}, c_{m'n'} \) the probability that \( \text{gcd}(c_{mn}, c_{m'n'}) = 1 \) is at least \( 1/4 \) 4. Thus only \( O(1) \) different divisors are required. In practice, it will not be known which denominators are divisors so every pair \( d_{mn}, d_{m'n'} \) with \( \text{gcd}(c_{mn}, c_{m'n'}) = 1 \) must be tested.

The second modification is simply allowing for the output \( j \) of QPF being useless. Let \( s \) denote the probability that \( j = \lceil e^{2L}/r \rceil \) or \( \lfloor e^{2L}/r \rfloor \) for some \( 0 < c < r \) where \( \lfloor \cdot \rfloor, \lceil \cdot \rceil \) denote rounding down and up respectively. Such values of \( j \) will be called useful as the denominators of the associated convergents are guaranteed to include a divisor of \( r \). The period \( r \) sought can always be found provided \( O(1/s) \) runs of QFT can be performed.

To summarize, as each new value \( j_m \) is measured, the denominators \( d_{mn} \) less than \( 2^L \) of the convergents of the continued fraction expansion of \( j_m/2^{2L} \) are substituted into \( f(k) = m^k \mod N \) to determine whether any \( f(d_{mn}) = 1 \) which would imply that \( r = d_{mn} \). If not, every pair \( d_{mn}, d_{m'n'} \) with associated numerators \( c_{mn}, c_{m'n'} \) satisfying \( \text{gcd}(c_{mn}, c_{m'n'}) = 1 \) must be tested to see whether \( r = \text{lcm}(d_{mn}, d_{m'n'}) \). Note that as shown in Fig. 11 if \( r \) is even or \( m'/2^s \mod N = \pm 1 \mod N \) then the entire process needs to be repeated \( O(1) \) times. Thus Shor’s algorithm always succeeds provided \( O(1/s) \) runs of QFT can be performed.

II. APPROXIMATE QUANTUM FOURIER TRANSFORM

A circuit that implements the QFT of Eq (4) is shown in Fig. 3a. Note the use of controlled rotations of magnitude \( \pi/2^d \). In matrix notation these 2-qubit operations correspond to

\[
\begin{pmatrix}
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & e^{i\pi/2^d}
\end{pmatrix}.
\]

The approximate QFT (AQFT) circuit is very similar with just the deletion of rotation gates with \( d \) greater than some \( d_{\text{max}} \). For example, Fig. 3b shows an AQFT with \( d_{\text{max}} = 1 \). Let \( [j]_m \) denote the \( m \)th bit of \( j \). The AQFT equivalent to Eq (4) is

\[
|k\rangle \rightarrow \frac{1}{\sqrt{2^{2L}}} \sum_{j=0}^{2^{2L}-1} |j\rangle \exp\left(\frac{2\pi i}{2^L} \sum_{mn} [j]_m [k]_n 2^{m+n}\right)
\]
where \( \sum_{mn} \) denotes a sum over all \( m, n \) such that \( 0 \leq m, n < 2L \) and \( 2L - d_{\text{max}} + 1 \leq m + n < 2L \). It has been shown by Coppersmith that the AQFT is a good approximation of the QFT in the sense that the phase of individual computational basis states in the output of the AQFT differ in angle from those in the output of the QFT by at most \( 2\pi L 2^{-d_{\text{max}}} \). The purpose of this paper is to investigate in detail the effect of using the AQFT in Shor’s algorithm.

**III. FAULT-TOLERANT CONSTRUCTION OF SMALL ANGLE ROTATION GATES**

When the 7-qubit Steane code \( [4] [8] [10] \) and its concatenated generalizations are used to do computation, only the limited set of gates CNOT, Hadamard (H), X, Z, S and \( S^\dagger \) can be implemented easily, where

\[
S = \begin{pmatrix} 1 & 0 \\ 0 & i \end{pmatrix}.
\]

Complicated circuits of depth in the hundreds and requiring a minimum of 22 qubits are required to implement the \( T \) and \( T^\dagger \) gates \( [4] \)

\[
T = \begin{pmatrix} 1 & 0 \\ 0 & e^{i\pi/4} \end{pmatrix}.
\]

The rationale of Eq (13) is that if \( U \) and \( V \) are similar, \( U^\dagger V \) will be close to the identity matrix (possibly up to some global phase) and the absolute value of the trace will be close to 2. By subtracting this absolute value from 2 and dividing by 2 a number between 0 and 1 is obtained. The overall square root is required to ensure that the triangle inequality

\[
dist(U, W) \leq dist(U, V) + dist(V, W)
\]

is satisfied.

The identity matrix is a good approximation of \( R_{128} \) in the sense that \( \text{dist}(R_{128}, I) = 8.7 \times 10^{-3} \). Eq (13) is only slightly better with \( \text{dist}(R_{128}, U_{31}) = 8.1 \times 10^{-3} \). A 46 gate sequence has been found satisfying \( \text{dist}(R_{128}, U_{46}) = 7.5 \times 10^{-4} \). Note that this is still only 10 times better than doing nothing. Further investigation of the properties of fault-tolerant approximations of arbitrary single-qubit unitaries will be performed in the near future. For the present discussion it suffices to know that the number of gates grows somewhere between linearly and quadratically with \( \ln(1/\delta) \) where \( \delta = \text{dist}(R, U) \), \( R \) is the rotation being approximated, and \( U \) is the approximating product of fault-tolerant gates (the exact scaling is not known). In particular, this means that approximating a rotation gate \( R_{2\delta} \) with accuracy \( \delta = 1/2^d \) requires a number of gates that grows linearly or quadratically with \( d \).

In addition to the inconveniently large number of fault-tolerant gates \( n_\delta \) required to achieve a given approximation \( \delta \), each individual gate in the approximating sequence must be implemented with probability of error \( p \) less than \( O(\delta/n_\delta) \). Note that \( \delta \) is not a probability of error but rather a measure of the distance between the ideal gate and the approximating product so this relationship is not exact. If the required probability \( p \sim \delta/n_\delta = 1/(n_\delta 2^d) \) is too small to be achieved using a single level of QEC, the technique of concatenated QEC must be used. Roughly speaking, if a given gate can be implemented with probability of error \( p \), adding an additional level of concatenation \( [12] \) leads to an error rate of \( cp^2 \) where \( c < 1/p \). If the Steane code is used with seven qubits for the code and an additional five qubits for fault-tolerant correction, every additional level of concatenation requires 12 times as many qubits. This implies that if a gate is to be implemented with accuracy \( 1/(n_\delta 2^d) \), the number of qubits \( q \) scales as \( O(d^{n_\delta 2^d}) = O(d^{1.58}) \). While this is a polynomial number of qubits, for even
moderate values of \(d\) this leads to thousands of qubits being used to achieve the required gate accuracy.

Given the complexity of implementing \(T\) and \(T^\dagger\) gates, the number of fault-tolerant gates required to achieve good approximations of arbitrary rotation gates and the large number of qubits required to achieve sufficiently reliable operation, it is clear that for practical reasons the use of \(\pi/2^d\) rotations must be restricted to those with very small \(d\).

IV. DEPENDENCE OF OUTPUT RELIABILITY ON PERIOD OF \(f(k) = m^k \text{mod} N\)

Different values of \(r\) (the period of \(f(k) = x^k \text{mod} N\)) imply different probabilities \(s\) that the value \(j\) measured at the end of QPF will be useful (see Fig 1). In particular, as discussed in Section II if \(r\) is a power of 2 the probability of useful output is much higher (see Fig 2). This section investigates how sensitive \(s\) is to variations in \(r\). Recall Eq (6) for the probability of measuring a given value of \(j\). When the AQFT (Eq (10)) is used this becomes

\[
Pr(j, r, L, d_{\text{max}}) = \left| \frac{\sqrt{r}}{2^L} \sum_{p=0}^{2^L/r - 1} \exp\left( \frac{2\pi i}{2^L} \sum_{m} [j]_m [p]_n 2^{m+n} \right) \right|^2
\]

(16)

The probability \(s\) of useful output is thus

\[
s(r, L, d_{\text{max}}) = \sum_{\{\text{useful } j\}} Pr(j, r, L, d_{\text{max}})
\]

(17)

where \(\{\text{useful } j\}\) denotes all \(j = \lfloor c 2^L/r \rfloor\) or \(\lceil c 2^L/r \rceil\) such that \(0 < c < r\). Fig 4 shows \(s\) for \(r\) ranging from 2 to \(2^L - 1\) and for various values of \(L\) and \(d_{\text{max}}\). The decrease in \(s\) for small values of \(r\) is more a result of the definition of \(\{\text{useful } j\}\) than an indication of poor data. When \(r\) is small there are few useful values of \(j \approx c 2^L/r\), \(0 < c < r\) and a large range states likely to be observed around each one resulting superficially in a low probability of useful output \(s\) as \(s\) is the sum of the probabilities of observing only values \(j = \lfloor c 2^L/r \rfloor\) or \(\lceil c 2^L/r \rceil\), \(0 < c < r\). However, in practice values much further from \(j \approx c 2^L/r\) can be used to obtain useful output. For example if \(r = 4\) and \(j = 16400\) the correct output value (4) can still be determined from the continued fraction expansion of 16400/65536 which is far from the ideal case of 16384/65536. To simplify subsequent analysis each pair \((L, d_{\text{max}})\) will from now on be associated with \(s(2^{L-1} + 2, L, d_{\text{max}})\) which corresponds to the minimum value of \(s\) to the right of the central peak. The choice of this point as a meaningful characterization of the entire graph is justified by the discussion above.

For completeness, Fig 5(e) shows the case of noisy controlled rotation gates of the form

\[
\begin{pmatrix}
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & e^{i(\pi/2^d + \delta)}
\end{pmatrix}.
\]

(18)

where \(\delta\) is a normally distributed random variable of standard deviation \(\sigma\). This has been included to simulate the effect of using approximate rotation gates built out of a finite number of fault-tolerant gates. The general form and probability of successful output can be seen to be similar despite \(\sigma = \pi/32\). This \(\sigma\) corresponds to \(\pi/2^{d_{\text{max}} + 2}\). For a controlled \(\pi/4\) rotation, single-qubit rotations of angle \(\pi/128\) are required, as shown in Fig 6. Fig 5(e) implies that it is acceptable for this rotation to be implemented within \(\pi/512\), implying

\[
U = \begin{pmatrix}
1 & 0 \\
0 & e^{i(\pi/128 + \pi/512)}
\end{pmatrix}
\]

(19)

is an acceptable approximation of \(R_{128}\). Given that \(\text{dist}(R_{128}, U) = 2.1 \times 10^{-3}\), the 46 fault-tolerant gate approximation of \(R_{128}\) mentioned above is adequate.

V. DEPENDENCE OF OUTPUT RELIABILITY ON INTEGER LENGTH AND GATE RESTRICTIONS

In order to determine how the probability of useful output \(s\) depends on both the integer length \(L\) and the minimum allowed controlled rotation \(\pi/2^{d_{\text{max}}}\), Eq (17) was solved with \(r = 2^{L-1} + 2\) as discussed in Section IV. Fig 6 contains semilog plots of \(s\) versus \(L\) for different values of \(d_{\text{max}}\). Note that Eq (17) grows exponentially more difficult to solve as \(L\) increases.

For \(d_{\text{max}}\) from 0 to 5, the exponential decrease of \(s\) with increasing \(L\) is clear. Asymptotic lines of best fit of the form

\[
s \propto 2^{-L/t}
\]

(20)

have been shown. Note that for \(d_{\text{max}} > 0\), the value of \(t\) increases by greater than a factor of 4 when \(d_{\text{max}}\) increases by 1. This enables one to generalize Eq (20) to an asymptotic lower bound valid for all \(d_{\text{max}} > 0\)

\[
s \propto 2^{-L/4^{d_{\text{max}} - 1}}
\]

(21)

with the constant of proportionality approximately equal to 1.

Keeping in mind that the required number of repetitions of QPF is \(O(1/s)\), one can relate \(L_{\text{max}}\) to \(d_{\text{max}}\) by introducing an additional parameter \(f_{\text{max}}\) characterizing the acceptable number of repetitions of QPF

\[
L_{\text{max}} \simeq 4^{d_{\text{max}} - 1} \log_2 f_{\text{max}}.
\]

(22)

Available RSA [15] encryption programs such as PGP typically use integers of length \(L\) up to 4096. The circuit
in $3$ runs in $150L^3$ steps when an architecture that can interact arbitrary pairs of qubits in parallel is assumed and fault-tolerant gates are used. Note that to first order in $L$ the number of steps does not increase as additional levels of QEC are used. Thus $\sim 10^{13}$ steps are required to perform a single run of QPF. On an electron spin or charge quantum computer running at 10GHz this corresponds to $\sim 15$ minutes of computing. If we assume $\sim 24$ hours of computing is acceptable then $f_{\text{max}} \sim 10^2$. Substituting these values of $L_{\text{max}}$ and $f_{\text{max}}$ into Eq. (22) gives $d_{\text{max}} = 6$ after rounding up. Thus provided controlled $\pi/64$ rotations can be implemented accurately, implying the need to accurately implement $\pi/128$ single-qubit rotations, it is conceivable that a quantum computer could one day be used to break a 4096-bit RSA encryption in a single day.

VI. CONCLUSION

We have demonstrated the robustness of Shor’s algorithm when a limited set of rotation gates is used. The length $L_{\text{max}}$ of the longest factorable integer can be related to the maximum acceptable runs of quantum period finding $f_{\text{max}}$ and the smallest accurately implementable controlled rotation gate $\pi/2^{d_{\text{max}}}$ via $L_{\text{max}} \sim 4^{d_{\text{max}} - 1} \log_2 f_{\text{max}}$. Integers thousands of digits in length can be factored provided controlled $\pi/64$ rotations can be implemented with rotation angle accurate to $\pi/256$. Sufficiently accurate fault-tolerant constructions of such controlled rotation gates have been described.
FIG. 6: Dependence of the probability of useful output from the quantum part of Shor’s algorithm on the length $L$ of the integer being factored for different levels of restriction of controlled rotation gates of angle $\pi/2^{d_{\text{max}}}$.

[1] P. W. Shor, Proceedings. 35th Annual Symposium on Foundations of Computer Science pp. 124–134 (1994).
[2] P. W. Shor, quant-ph/0005003 (2000).
[3] D. Coppersmith, IBM Research Report p. RC19642 (1994).
[4] M. A. Nielson and I. L. Chuang, Quantum Computation and Quantum Information (Cambridge University Press, Cambridge, 2000).
[5] V. Vedral et al., Phys. Rev. A 54, 147 (1996).
[6] P. Gossett, quant-ph/9808061 (1998).
[7] S. Beauregard, Quantum Inform. Compu. 3, 175 (2003).
[8] C. Zalka, quant-ph/9806084 (1998).
[9] A. M. Steane, Proc. R. Soc. London A 425, 2551 (1996).
[10] A. R. Calderbank and P. W. Shor., Phys. Rev. A 54, 1098 (1995).
[11] A. Y. Kitaev, Russ. Math. Surv. 52, 1191 (1997).
[12] E. Knill et al., quant-ph/9610011 (1996).
[13] R. L. Rivest et al., Communications of the ACM 21, 120 (1978).
[14] G. Burkard et al., Fortschr. Physik 48, 965 (2000).
[15] L. C. L. Hollenberg et al., cond-mat/0306235 (2003).