SCIENCE CHINA
Information Sciences

• POSITION PAPER •

July 2024, Vol. 67, Iss. 7, 173501:1–173501:16
https://doi.org/10.1007/s11432-023-3961-3

Shor's algorithm does not factor large integers in the presence of noise

Jin-Yi CAI

College of Letters and Science, University of Wisconsin, Madison WI 53706, *USA*

Received 18 October 2023/Revised 9 February 2024/Accepted 21 February 2024/Published online 18 June 2024

Abstract We consider Shor's quantum factoring algorithm in the setting of noisy quantum gates. Under a generic model of random noise for (controlled) rotation gates, we prove that the algorithm does not factor integers of the form pq when the noise exceeds a vanishingly small level in terms of n—the number of bits of the integer to be factored, where p and q are from a well-defined set of primes of positive density. We further prove that with probability $1 - o(1)$ over random prime pairs (p, q), Shor's factoring algorithm does not factor numbers of the form pq, with the same level of random noise present.

Keywords random noise, Shor's algorithm, rotation gates, quantum computing, prime factorization

1 Introduction

One of the most stunning achievements of computer science in the last several decades is Shor's quantum algorithm to factor large integers [1, 2]. The algorithm can provably factor an n-bit integer in polynomial time with high probability, assuming certain quantum operations can be performed. These are called quantum logic gates. In particular, they include the familiar Hadamard gate $H = \frac{1}{\sqrt{2}}\left[\begin{smallmatrix}1 & 1\\ 1 & -1\end{smallmatrix}\right]$, the rotation gates (Phase) $S = \left[\begin{smallmatrix}1 & 0\\ 0 & i\end{smallmatrix}\right]$, ($\pi/8$ gate) $T = \left[\begin{smallmatrix}1 & 0\\ 0 & \mathrm{e}^{2\pi\mathrm{i}/8}\end{smallmatrix}\right]$, and more generally $R_k = \left[\begin{smallmatrix}1 & 0\\ 0 & \mathrm{e}^{2\pi\mathrm{i}/2^k}\end{smallmatrix}\right]$, and their controlled versions. Note that $S = R_2$ and $T = R_3$.

It has often been pointed out that the availability of these quantum gates at high precision (with arbitrarily small angles in R_k with $k \to \infty$) is a challenge, both theoretically in terms of the limit of physical theory and practically on engineering grounds [3–6][1][2]. To a large extent, such concerns motivated another great intellectual achievement that is the development of quantum error correcting codes [7–11]. There is a substantial body of work on fault tolerant quantum computing, starting with Shor's work [12]. Strong threshold theorems are proven which show that in certain error models, if the error rate is below a certain threshold, quantum computation can achieve, at least theoretically, arbitrarily high accuracy [10, 13–18]. These are beautiful mathematical theorems. But fundamentally they assume that the group U(2) (or SU(2) if we factor out an irrelevant phase factor) exactly corresponds to operations on a qubit in reality, especially in its composition—that group composition, in its infinite precision defined over $\mathbb{C}$, exactly corresponds to sequential application of realizable physical quantum operations. Opinions differ, as to whether such arbitrary precision is ever achievable. It is certainly a possibility. However, this author is skeptical about this, based on the belief that quantum mechanics itself (just as any other physical theory) is not, and is not meant to be, infinitely accurate when describing reality (some speculative comments are in Section 5). Meanwhile, enormous efforts have been underway in the past few decades, with much renewed momentum and enthusiasm more recently, and with the goal of achieving ever more accurate hardware implementations of quantum circuitry.

In this paper, we consider Shor's quantum factoring algorithm in the setting where each quantum controlled rotation gate is subject to a small random noise in the angle. We assume each application of

Email: jyc@cs.wisc.edu

[1] Gil Kalai. https://gilkalai.wordpress.com/2022/05/26/waging-war-on-quantum/.

[2] Leonid Levin. https://www.cs.bu.edu/fac/lnd/expo/qc.htm.

the controlled-R_k gate is given an independent random error of angle $\mathrm{e}^{2\pi \mathrm{i}\epsilon r/2^k}$. Thus, when the control bit is 1, the operator R_k is substituted by $\widetilde{R_k} = \left[\begin{smallmatrix}1 & 0\\ 0 & \mathrm{e}^{2\pi\mathrm{i}(1+\epsilon r)/2^k}\end{smallmatrix}\right]$, where r is an independent noise random variable distributed $r \sim N(0,1)$, and ϵ is a global magnitude parameter. So, the controlled-$\widetilde{R_k}$ gate is

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & \rho_k \xi_k \end{bmatrix},$$

where $\rho_k = \rho_{k,\epsilon} = \mathrm{e}^{2\pi\mathrm{i}\epsilon r/2^k}$ and $\xi_k = \mathrm{e}^{2\pi\mathrm{i}/2^k}$. We show that there exist positive constants $c, c' > 0$ such that if $\epsilon > cn^{-1/3}$, then Shor's algorithm does not factor n-bit integers of the form pq, where p and q are from a well-defined set of primes of density $> c'$. This is the first provable statement of such failure of Shor's algorithm under any error model.

The noise model is similar to that of [19] (see also [20–22]). The specific random noise model including the independent normal distribution picked in this paper is not essential, as the proof will clearly show, but it is chosen to present the essential idea of the proof most transparently. For example, the noise r.v. r being distributed $\sim N(0,1)$ (standard normal distribution) can be replaced by any reasonable alternative distribution such as uniform $U[-1,1]$ or uniform bits from $\{-1,0,1\}$. While each individual controlled-R_k gate is assumed to be accompanied by an independent r.v. r for noise, when an individual controlled-R_k gate is applied, the same randomly perturbed controlled-R_k gate is applied to each term in a sum of superpositions of quantum states. Regarding the random noise model, we do not make any claim that this model accurately reflects "reality"; our purpose is only to show that some vanishing amount of noise can already provably destroy the algorithm. Unruh [5] argued that quantum calculations require coherence to be maintained during the course of the calculation (however no explicit theorem is stated). Landauer [4] had emphasized the need to examine effect of imperfections of realistic quantum devices.

An important modification of Shor's algorithm by Coppersmith [23] shows that if we just ignore (not to perform) all (controlled-) R_k-gates for sufficiently large $k \geqslant b$, where b is some global parameter, then Shor's algorithm still retains its effectiveness (and uses a reduced number of quantum gates). The specific suggested change [23] for 500-qubits, which would require rotations of magnitude $2\pi/2^{500}$ in Shor's original algorithm, is to ignore all rotations of angle smaller than $2\pi/2^{20}$. It is estimated that this would incur an error on the order of 1% in the probability of each desirable final state. Asymptotically, Coppersmith improved the precision requirement of exponentially small angles to just slightly less than π/n. This is of enormous practical implications. This version of Shor's algorithm is called the "banded" version with parameter b, which is set to be slightly greater than $\log n$, rather than n in the original version. Nonetheless, rotation gates (as primitive steps of the algorithm) of asymptotically infinitely small angles would still be required as n, the number of bits to be factored, tends to infinity.

Our result is consistent with Coppersmith's improvement. Indeed we will present our proof in the "banded" version, with perfect controlled-R_k-gates for all $k < b$, but every controlled-R_k-gate is replaced by a controlled-$\widetilde{R_k}$-gate for all $k \geqslant b$; i.e., it is independently perturbed by a random noise. Our negative result will be stated in terms of $b + \log_2(1/\epsilon)$. When $b + \log_2(1/\epsilon) < \frac{1}{3}\log_2 n - c$ for some constant $c > 0$, the noise takes hold so as to destroy the desired peak in the probability of observing a useful state that leads to factorization. This condition is essentially equivalent to having both b being less than a small constant multiple of $\log n$ and ϵ greater than the reciprocal of a small positive power of n. Note that the statement for the banded version is a stronger result, in the sense that the unbanded version where controlled-$\widetilde{R_k}$-gate is used for all k is an easy consequence. We prove that, if $b + \log_2(1/\epsilon) < \frac{1}{3}\log_2 n - c$, Shor's algorithm does not factor n-bit integers of the form pq, where p and q are from a well-defined set of primes of positive density $c' > 0$[3]. The proof will in fact show that the same failure happens under the same condition, even if the noise gates are applied only at the single level R_b, with all other controlled-R_k-gates applied perfectly for $k \neq b$ (or alternatively, no controlled-R_k-gates are applied at all for $k > b$ as in the banded version by Coppersmith).

[3] We note that the results from [19–22] are generally stated in the opposite direction. Under plausible, but ultimately heuristic, assumptions for the behavior of various sums, augmented by numerical simulations, they suggest that if b is not too large compared to n, Shor's algorithm can tolerate imprecisions of rotation angles. Some small concrete values of n are on the order of 10 qubits ($n = 10, 14$). These values are quite outside the range where our proof applies. Their numerical simulation does seem to suggest a logarithmic threshold of b. Thus, these positive results are not logically inconsistent with, and in fact, complement our proof. Please note the notation b in [21] is our $b - 2$.

Theorem 1. There exist constants $c, c' > 0$, such that if each controlled-R_k-gate in the quantum Fourier transform circuit is replaced by controlled-$\widetilde{R_k}$-gate for all $k \geqslant b$, where $b + \log_2 (1/\epsilon) < \frac{1}{3} \log_2 n - c$, then with exponentially small exceptional probability, Shor's algorithm does not factor n-bit integers of the form pq, where p and q are from a well-defined set of primes of density $> c'$.

Here "exceptional probability" is over the random choices of Shor's algorithm as well as probabilistic outcomes of quantum measurements. More precisely, the expectation over random noise r's, of the success probability (over the random choices of the algorithm and quantum measurements) of the algorithm is exponentially small in n. This will be the meaning of "does not factor" below.

Theorem 2. If $b + \log_2 (1/\epsilon) < \frac{1}{3} \log_2 n - c$, then the statement in Theorem 1 still holds, if only each controlled-R_b-gate is replaced by a controlled-$\widetilde{R_b}$-gate while all other controlled-R_k-gates remain unchanged. Alternatively, the same statement holds if each controlled-R_k-gate is (1) applied perfectly for $k < b$, (2) replaced by a controlled-$\widetilde{R_b}$-gate for $k = b$, and (3) deleted for $k > b$.

Our proof focuses on the essential "period-finding" part of Shor's algorithm that uses the quantum Fourier transform (QFT). In our proof, we use a theorem of Fouvry [24]. This theorem states that the set of all primes p such that the largest prime factor in $p - 1$ is greater than $p^{2/3}$ has positive density among all primes. We use this theorem to produce candidate inputs of the form $N = pq$ to Shor's algorithm where p and q are of this type, and argue that a random element $x \in \mathbb{Z}_N^*$ has (exponentially) large order $\omega = \omega_N(x)$ as an element of the multiplicative group $\mathbb{Z}_N^*$. This large order ω allows us to give a lower bound for a lattice counting argument, which leads to a sufficiently large number of independent perturbations in the complex arguments (in the exponent) in a crucial sum of exponentials, (which would have been a perfect geometric sum without noise) in the analysis of Shor's algorithm. This perturbation, at the appropriate setting of parameters, destroys this geometric sum, and degrades the probability of observing any useful quantum state to negligible, and thus fails to gain any useful information on the period ω.

Our proof can be adapted to more general primes beyond those guaranteed by Fouvry's theorem.

Theorem 3. There exists a constant $c > 0$, such that for random primes p and q chosen uniformly from all primes of binary length m, if $b + \log_2 (1/\epsilon) < \frac{1}{3} \log_2 m - c$, as $m \to \infty$ with probability $1 - o(1)$, Shor's algorithm with noisy rotation gates does not factor $N = pq$. A version analogous to Theorem 2 also holds for random primes.

We make a few brief remarks. Arguably, factoring integers $N = pq$ for random primes p and q is more important in cryptography than for primes that satisfy the property in Fouvry's theorem, and the statement of failure probability being $1 - o(1)$ is stronger than that of positive density guaranteed by Fouvry's theorem. We first present the proof for the latter, and relegate the proof of Theorem 3 to Section 4, to demonstrate the main idea of our proof in a simple setting of how random noise degrades the performance of Shor's algorithm. The additional work needed for Theorem 3 is mainly of a number theoretic nature.

One can further prove other versions of Theorem 3. For example, we can restrict the random primes p and q to be of length m and both $\equiv 3 \bmod 4$, so that the numbers $N = pq$ are the so-called Blum integers, which are favored in cryptography [25]. In this paper we do not present these generalizations. Despite the strong failure demonstrated by the proof, our theorems do not rule out the possibility that at some future time, some quantum algorithm may prove superior to the best "classical" factoring algorithms for factoring integers of a certain size, in practice. But our proof indicates that there is a limit to this possible superiority when n is large, if arbitrarily small random noise cannot be eliminated.

Section 2 gives some preliminaries. Section 3 presents the proof of Theorems 1 and 2. Section 4 presents the proof of Theorem 3. In Section 5 some speculative comments are presented. The Church-Turing thesis identifies computability with Turing machine computability. The Strong Church-Turing thesis identifies feasible computability with (probabilistic) polynomial-time computability, namely P or BPP. Many people have made strong arguments [11] supporting the viewpoint that Shor's algorithm presents a convincing evidence that this Strong Church-Turing thesis should be modified so that P or BPP is replaced by BQP. This author is personally not convinced of this, and makes some speculative comments on that. These comments should not be conflated with the theorems proven in the paper.

2 Preliminaries

Fouvry's theorem. Let $N = pq$, where p and q are distinct odd primes. By the Chinese remainder theorem, the multiplicative group $\mathbb{Z}_N^* = \{m \in \mathbb{Z}_N \mid \gcd(m, N) = 1\}$ (invertible elements in $\mathbb{Z}_N$) is isomorphic to the direct product $\mathbb{Z}_p^* \times \mathbb{Z}_q^*$. Moreover, $\mathbb{Z}_p^*$ is a cyclic group of order $p-1$, and is isomorphic to a direct product of factors according to the prime factorization of $p-1$; and similarly for $\mathbb{Z}_q^*$. If $p-1 = 2^e p_1^{e_1} \cdots p_k^{e_k}$, where $p_1 < \cdots < p_k$ are distinct odd primes, then $\mathbb{Z}_p^*$ is isomorphic to $\mathbb{Z}_{2^e} \times \mathbb{Z}_{p_1^{e_1}} \times \cdots \times \mathbb{Z}_{p_k^{e_k}}$. Let $P^+(m)$ denote the largest prime in the prime factorization of m.

Theorem 4 (Fouvry). There exist constants $c > 0$ and $n_0 > 0$, such that for all $x > n_0$,

$$|\{p \mid p \text{ is a prime}, \ p < x, \ \text{and } P^+(p-1) > p^{2/3}\}| \geqslant c\frac{x}{\log x}.$$

We say a prime p satisfies the Fouvry property if $P^+(p-1) > p^{2/3}$. If $N = pq$, where p and q are distinct odd primes satisfying the Fouvry property, then clearly $p' = P^+(p-1)$ appears with exponent 1 in the factorization of $p-1$, and so does $P^+(q-1)$ in the factorization of $q-1$. If $p' = P^+(p-1) > P^+(q-1)$, then $\mathbb{Z}_{p'}$ appears as an isolated factor in the direct product form of $\mathbb{Z}_N^*$. Thus, with probability $\geqslant 1 - 1/p' > 1 - \frac{1}{\max\{p^{2/3}, q^{2/3}\}} \geqslant 1 - N^{-1/3}$, a random element x in $\mathbb{Z}_N^*$ has order at least $p' > \max\{p^{2/3}, q^{2/3}\} \geqslant N^{1/3}$. If it so happens that $p' = P^+(p-1) = P^+(q-1)$, then $\mathbb{Z}_{p'} \times \mathbb{Z}_{p'}$ appears as a factor in the direct product form of $\mathbb{Z}_N^*$. In this case, a random element x in $\mathbb{Z}_N^*$ also has order at least $p' > N^{1/3}$ with probability $\geqslant 1 - 1/(p')^2 \geqslant 1 - N^{-2/3}$. Thus, in either case, in terms of the number of bits, such products $N = pq$ have the property that a random element x in $\mathbb{Z}_N^*$ has an exponentially large period, $\omega = \omega_N(x) \geqslant \max\{P^+(p-1), P^+(q-1)\} > N^{1/3}$, with exponentially small exceptional probability. Below we assume ω has this property.

Let $\mathrm{ord}_2(x)$ denote the highest power of 2 that divides x. If $e = \mathrm{ord}_2(p-1)$ and $e' = \mathrm{ord}_2(q-1)$, then we have $2^e < (P^+(p-1))^{1/2}$ and $2^{e'} < (P^+(q-1))^{1/2}$, and thus $\omega = \omega_N(x)$ satisfies $\mathrm{ord}_2(\omega) \leqslant \max\{e, e'\} < \frac{\log_2 \omega}{2}$, for any $x \in \mathbb{Z}_N^*$. We conclude the following.

Lemma 1. Let p and q be distinct odd primes satisfying the Fouvry property, and let $N = pq$, then over a random $x \in \mathbb{Z}_N^*$,

$$\mathrm{Pr.}\left(\omega_N(x) > N^{1/3} \quad \text{and} \quad \mathrm{ord}_2(\omega_N(x)) < \frac{\log_2 \omega_N(x)}{2}\right) > 1 - \frac{1}{N^{1/3}}.$$

Sum of random unit vectors. Let $\xi_m = \mathrm{e}^{2\pi \mathrm{i}/m}$ be a primitive root of unity of order m. Let $X_i \sim N(0,1)$, $i = 1, 2, \ldots, n$, be a finite sequence of independent and identically distributed (i.i.d.) normal random variables. Let $\{S_k \subseteq [n] \mid 1 \leqslant k \leqslant K\}$ be a finite collection of sets such that each pairwise symmetric difference $S_j \Delta S_k$ has cardinality $\geqslant m^2 t$, for all $j \neq k$. Let $\Sigma_k = \sum_{i \in S_k} X_i$ be the sum of X_i for $i \in S_k$. We will give a simple estimate for the expectation of

$$|\xi_m^{\Sigma_1} + \xi_m^{\Sigma_2} + \cdots + \xi_m^{\Sigma_K}|^2. \tag{1}$$

Expanding the square norm expression we get

$$K + \sum_{1 \leqslant j < k \leqslant K} (\xi_m^{\Sigma_j - \Sigma_k} + \xi_m^{\Sigma_k - \Sigma_j}) = K + 2 \sum_{1 \leqslant j < k \leqslant K} \cos\left((\Sigma_j - \Sigma_k)\frac{2\pi}{m}\right).$$

Let $T_{jk} = (\Sigma_j - \Sigma_k)\frac{2\pi}{m}$. Note that $\Sigma_j - \Sigma_k = \sum_{i \in S_j \Delta S_k} (\pm X_i)$ is a sum of at least $m^2 t$ distinct (thus independent) r.v. $\pm X_i$ distributed i.i.d. $\sim N(0,1)$. Therefore, each T_{jk} is a random variable normally distributed $\sim N(0, \sigma_{jk}^2)$, with standard deviation $\sigma_{jk} = \sqrt{|S_j \Delta S_k|} \cdot \frac{2\pi}{m} \geqslant 2\pi\sqrt{t}$.

Moments of even orders of a normal random variable $Y \sim N(0, \sigma^2)$ are known as follows [26]:

$$\mathbb{E}[Y^{2k}] = \sigma^{2k}(2k-1)!!,$$

where $\mathbb{E}$ denotes expectation, from which we get (by the dominated convergence theorem, the exchange of orders of summation and integration is justified)

$$\begin{aligned}\mathbb{E}[\cos(T_{jk})] &= 1 - \frac{\sigma_{jk}^2}{2!}(2-1)!! + \frac{\sigma_{jk}^4}{4!}(4-1)!! - \frac{\sigma_{jk}^6}{6!}(6-1)!! + \cdots \\ &= \mathrm{e}^{-\sigma_{jk}^2/2}\end{aligned}$$

$$\leqslant \mathrm{e}^{-2\pi^2 t}.$$

Hence, the expectation of (1) is at most $K + 2\binom{K}{2}\mathrm{e}^{-2\pi^2 t}$.

We will need a slight generalization of this. Let $\sigma > 0$, and let $\varphi_k \in [0, 2\pi)$ be any angle, $1 \leqslant k \leqslant K$. We replace each Σ_k by $\varphi_k + \sigma \sum_{i \in S_k} X_i$. Then, we have the following.

Lemma 2. Let $\sigma > 0$ and $\xi_m = \mathrm{e}^{2\pi \mathrm{i}/m}$. Let $X_i \sim N(0, 1)$, i.i.d. for $i = 1, 2, \ldots, n$, and let $\{S_k \subseteq [n] \mid 1 \leqslant k \leqslant K\}$ be a finite collection of sets. Assume all except at most δ fraction of pairwise symmetric differences $S_j \Delta S_k$ have cardinality $\geqslant (m/\sigma)^2 t$ for $j \neq k$. Let $\Sigma_k = \varphi_k + \sigma \sum_{i \in S_k} X_i$, where $\varphi_k \in [0, 2\pi)$. Then,

$$\mathbb{E}[|\xi_m^{\Sigma_1} + \xi_m^{\Sigma_2} + \cdots + \xi_m^{\Sigma_K}|^2] \leqslant K + 2\delta \binom{K}{2} + 2(1 - \delta)\binom{K}{2}\mathrm{e}^{-2\pi^2 t}.$$

Proof. Let $T_{jk} = \frac{2\pi\sigma}{m}(\sum_{i \in S_j} X_i - \sum_{i \in S_k} X_i)$. We only need to note in addition to the above that

$$\cos(\varphi + T_{jk}) = \cos\varphi \cos T_{jk} - \sin\varphi \sin T_{jk},$$

and we have $\cos\varphi \leqslant 1$ for any φ, and $\mathbb{E}[\sin T_{jk}] = 0$ since sin is an odd function and T_{jk} is symmetrically distributed. The lemma follows.

3 Corrupted geometric sums

Suppose N is an integer we wish to factor, and $2^n \approx N^2$ as in [2][4]. For definiteness assume $2^{n-1} < N^2 \leqslant 2^n$. Assume ω is the period of the function $f(k) = x^k \bmod N$ for a randomly chosen $x \in \mathbb{Z}_N^*$, and by Lemma 1 we assume $\omega > N^{1/3}$ and $\mathrm{ord}_2(\omega) < \frac{\log_2 \omega}{2}$. Also $\omega < N$ clearly.

Let us write out a few terms as the controlled-R_k gates are applied successively in the QFT circuit (see p.219 of [11]), but now with random noise added whenever the controlled rotation gate is R_k-gates with $k \geqslant b$, i.e., we apply controlled-R_k-gates when $k < b$ but controlled-$\widetilde{R_k}$-gates for all $k \geqslant b$. As the first controlled-R_k-gate has $k = 2$, we may assume $b > 1$. Suppose we start with the state $|u\rangle = |u_{n-1} \cdots u_1 u_0\rangle$. After the first gate H on the qubit $|u_{n-1}\rangle$, we have the state

$$\frac{1}{2^{1/2}}\left(|0\rangle + \mathrm{e}^{2\pi \mathrm{i}\ 0.u_{n-1}}|1\rangle\right)|u_{n-2} \cdots u_0\rangle.$$

The next is the controlled-R_2-gate on target qubit $|u_{n-2}\rangle$ controlled by the leftmost qubit (which was initially $|u_{n-1}\rangle$). If $b > 2$ then the unperturbed controlled-R_2-gate is applied, after which we have

$$\frac{1}{2^{1/2}}\left(|0\rangle + \mathrm{e}^{2\pi \mathrm{i}\ 0.u_{n-1}u_{n-2}}|1\rangle\right)|u_{n-2} \cdots u_0\rangle.$$

If $b = 2$ the above statement is vacuous, and the perturbed controlled-$\widetilde{R_2}$-gate is applied instead. The random noise starts at the controlled-R_b-gate, after which we get

$$\frac{1}{2^{1/2}}\left(|0\rangle + \mathrm{e}^{2\pi \mathrm{i}\ \left[0.u_{n-1} \cdots u_{n-b} + \frac{\epsilon}{2^b} u_{n-b} r_0^{(0)}\right]}|1\rangle\right)|u_{n-2} \cdots u_0\rangle,$$

where $r_0^{(0)} \sim N(0, 1)$.

After all the rotation gates controlled by the leftmost qubit (initially $|u_{n-1}\rangle$) we have

$$\frac{1}{2^{1/2}}\left(|0\rangle + \mathrm{e}^{2\pi \mathrm{i}\ [0.u_{n-1} \cdots u_0 + \frac{\epsilon}{2^b}(u_{n-b} r_0^{(0)} + \frac{u_{n-b-1} r_1^{(0)}}{2} + \cdots + \frac{u_0 r_{n-b}^{(0)}}{2^{n-b}})]}|1\rangle\right)|u_{n-2} \cdots u_0\rangle, \tag{2}$$

where $r_0^{(0)}, \ldots, r_{n-b}^{(0)}$ are i.i.d. $\sim N(0, 1)$.

Then, similarly, after all the rotation gates controlled by the two leftmost qubits (initially $|u_{n-1}u_{n-2}\rangle$), we have

$$\frac{1}{2^{2/2}}\left(|0\rangle + \mathrm{e}^{2\pi \mathrm{i}\ [0.u_{n-1} \cdots u_0 + \frac{\epsilon}{2^b}(u_{n-b} r_0^{(0)} + \cdots + \frac{u_0 r_{n-b}^{(0)}}{2^{n-b}})]}|1\rangle\right)$$

[4] Thus N has $\approx n/2$ bits, a slight change in notation from Section 1.

$$\otimes\left(|0\rangle + \mathrm{e}^{2\pi\mathrm{i}\ [0.u_{n-2}\cdots u_0+\frac{\epsilon}{2^b}(u_{n-b-1}r_0^{(1)}+\cdots+\frac{u_0 r_{n-b-1}^{(1)}}{2^{n-b-1}})]}|1\rangle\right)|u_{n-3}\cdots u_0\rangle, \tag{3}$$

where $r_0^{(0)},\ldots,r_{n-b}^{(0)},r_0^{(1)},\ldots,r_{n-b-1}^{(1)}$ are i.i.d. $\sim N(0,1)$.

The circuit continues to apply controlled rotation gates with random noise starting at the controlled-R_b-gate, producing a final expression with n tensor factors. When written out the tensor product, this is a sum indexed by $|v_{n-1}\cdots v_0\rangle$, such that $v_0 = 0$ or 1 corresponds to selecting respectively the term $|0\rangle$ or $\mathrm{e}^{2\pi\mathrm{i}\ [\cdots]}|1\rangle$ in (2) (or equivalently, to selecting one of the two terms in the first tensor factor in (3)), and $v_1 = 0$ or 1 corresponds to selecting respectively the term $|0\rangle$ or $\mathrm{e}^{2\pi\mathrm{i}\ [\cdots]}|1\rangle$ in the second tensor factor in (3), and similarly for $v_s = 0$ or 1, for all $0 \leqslant s \leqslant n-1$.

The crucial step in Shor's algorithm, after the quantum Fourier transform, is to take a quantum measurement, with the property that the probability of observing a state that is close to an integral multiple of $\frac{2^n}{\omega}$ is high. Such a state has an n-bit integer expression $v \in \{0,1\}^n$ that has value close to the rational number $\frac{2^n}{\omega}j$, for some $0 \leqslant j \leqslant \omega$. States $|v\rangle$ such that the number v is not close to an integral multiple of $\frac{2^n}{\omega}$ have a negligible probability of being observed, while states in a small vicinity of each of the integral multiples of $\frac{2^n}{\omega}$ get observed with probability on the order of $1/\omega$ (per each multiple), and these add up to give a good probability that some such state is observed, whereby the period is deduced with good probability. (This paper omits steps of the continued fraction algorithm in the post quantum processing steps.)

For each v, the probability of $|v\rangle$ being observed has an expression as a square norm of a sum over a set of the form $u \in \{u^* + k\omega : k \geqslant 0,\ \text{and}\ u^* + k\omega < 2^n\}$ (for some initial $0 \leqslant u^* < \omega$), with cardinality K, which is approximately $2^n/\omega$. For $u^{(k)} = u^* + k\omega$, we write the n-bit integers $u^{(k)} = \sum_{s=0}^{n-1} u_s^{(k)} 2^s$ and $v = \sum_{s=0}^{n-1} v_s 2^s$. When there is no noise in the controlled-R_k-gates used in the QFT, this probability expression for observing $|v\rangle = |v_{n-1}\ldots v_1 v_0\rangle$ can be written as

$$\frac{1}{2^n K}\left|\sum_{k=0}^{K-1}\exp\left\{2\pi\mathrm{i}\sum_{t=1}^{n}\frac{\sum_{s=0}^{n-t}u_{n-t-s}^{(k)}v_s}{2^t}\right\}\right|^2.$$

With independent random noise present starting with controlled-R_b-gates, this becomes

$$\frac{1}{2^n K}\left|\sum_{k=0}^{K-1}\exp\left\{2\pi\mathrm{i}\left[\sum_{t=1}^{n}\frac{\sum_{s=0}^{n-t}u_{n-t-s}^{(k)}v_s}{2^t}+\frac{\epsilon}{2^b}\left\{\left(u_{n-b}^{(k)}r_0^{(0)}+\cdots+\frac{u_0^{(k)}r_{n-b}^{(0)}}{2^{n-b}}\right)v_0\right.\right.\right.\right.$$
$$\left.\left.\left.\left.+\left(u_{n-b-1}^{(k)}r_0^{(1)}+\cdots+\frac{u_0^{(k)}r_{n-b-1}^{(1)}}{2^{n-b-1}}\right)v_1+\cdots+u_0^{(k)}r_0^{(n-b)}v_{n-b}\right\}\right]\right\}\right|^2, \tag{4}$$

where

$$r_0^{(0)},\ldots,r_{n-b}^{(0)},r_0^{(1)},\ldots,r_{n-b-1}^{(1)},\ldots,r_0^{(n-b-1)},r_1^{(n-b-1)},r_0^{(n-b)}$$

are random variables i.i.d. $\sim N(0,1)$.

Our first goal is to show that among states $|v\rangle$ such that the binary number v is close to an integral multiple $\frac{2^n}{\omega}j$ (for some $0 \leqslant j \leqslant \omega$), it is the case that for most j, a linear number of bits in the binary expansion of v are one: $v_s = 1$. This will leave us with a linear number of terms of the form in the exponent

$$\frac{2\pi\mathrm{i}\epsilon}{2^b}\left(u_{n-b-s}^{(k)}r_0^{(s)}+\frac{u_{n-b-s-1}^{(k)}r_1^{(s)}}{2}+\cdots+\frac{u_0^{(k)}r_{n-b-s}^{(s)}}{2^{n-b-s}}\right)v_s.$$

Eventually we will show that, fixing any such v, among those s where $v_s = 1$, for most k, there are a linear number of terms with $u_{n-b-s}^{(k)} = 1$, which will give us the perturbation as a sum of $\frac{2\pi\mathrm{i}\epsilon}{2^b}\cdot r_0^{(s)}$.

Let us consider integers $v = \lfloor\frac{2^n}{\omega}j\rfloor$, for $0 \leqslant j < \omega$; it will be clear from the proof below that what is proven is also true for any v in the vicinity of a polynomial range of such a number.

For $0 \leqslant j < \omega$, the integer $v = \lfloor\frac{2^n}{\omega}j\rfloor$ has the i-th leading bit $v_{n-i} = 1$ iff the i-th most significant bit, among the first n bits, in the binary expansion of $\frac{j}{\omega}$ is 1. This is true iff for some $1 \leqslant k \leqslant 2^{i-1}$,

$$\frac{2k-1}{2^i} \leqslant \frac{j}{\omega} < \frac{2k}{2^i},$$

which is equivalent to

$$(2k-1)\frac{\omega}{2^i} \leqslant j < 2k\frac{\omega}{2^i}. \tag{5}$$

So, j needs to be placed in the alternate ("odd" indexed) segments of length $\frac{\omega}{2^i}$. This is a lattice counting problem.

Recall that $\omega > N^{1/3} \approx 2^{n/6}$. We take $i_0 = \lfloor \frac{3}{4}\log_2 \omega \rfloor \geqslant \lfloor \frac{1}{4}\log_2 N \rfloor = \Omega(n)$. Then $\frac{\omega}{2^{i_0}} \geqslant \omega^{1/4} > N^{1/12} = 2^{\Omega(n)}$. We will only count those i-th (significant) bits v_{n-i} that are one, within $1 \leqslant i \leqslant i_0$, and first show that for most j, even just among the first i_0 bits $v_{n-1}, \ldots, v_{n-i_0}$, there are a linear number of ones. (Any additional bits that are 1 can only add more noise to the perturbation. But our proof does not depend on this fact.)

Now we divide the range $[0, \omega)$ of real numbers into 2^{i_0} segments of equal length $\frac{\omega}{2^{i_0}}$

$$I_\alpha = \left\{ x \in \mathbb{R} \mid \frac{\omega}{2^{i_0}}(\alpha)_2 \leqslant x < \frac{\omega}{2^{i_0}}((\alpha)_2 + 1) \right\},$$

where $\alpha \in \{0,1\}^{i_0}$ is a binary string, and $(\alpha)_2$ is the binary number it represents[5)].

Note that any real interval of the form $[A, A+B)$ has either $\lfloor B \rfloor$ or $\lfloor B \rfloor + 1$ many integers. Thus, each I_α contains either $\lfloor \frac{\omega}{2^{i_0}} \rfloor$ or $\lfloor \frac{\omega}{2^{i_0}} \rfloor + 1$ many integers, which is $\frac{\omega}{2^{i_0}} + \eta$ for some $-1 \leqslant \eta \leqslant 1$. We consider two distributions on the integers $0 \leqslant j < \omega$. Let Pr. denote the uniform distribution and let Pr_α denote the distribution induced by first picking $\alpha \in \{0,1\}^{i_0}$ uniformly, and then picking $j \in I_\alpha$ uniformly. They are exponentially close: for any $0 \leqslant j < \omega$, $\mathrm{Pr.}(j) = 1/\omega$, and

$$\mathrm{Pr}_\alpha(j) = \frac{1}{2^{i_0}} \frac{1}{\frac{\omega}{2^{i_0}} + \eta} = \frac{1}{\omega + \eta 2^{i_0}} = \frac{1}{\omega} \cdot \frac{1}{1 + \eta\frac{2^{i_0}}{\omega}} = \left(1 \pm 2^{-\Omega(n)}\right) \cdot \mathrm{Pr.}(j). \tag{6}$$

Let $\alpha = \alpha_1\alpha_2 \cdots \alpha_{i_0}$. Consider any $j \in I_\alpha$. If $\alpha_{i_0} = 1$ (i.e., $(\alpha)_2$ is odd), then j satisfies (5) for $i = i_0$. Now suppose $\alpha_{i_0-1} = 1$, then

$$\frac{\omega}{2^{i_0}}(\alpha_1 \cdots \alpha_{i_0-1}\alpha_{i_0})_2 \geqslant \frac{\omega}{2^{i_0}}(\alpha_1 \cdots \alpha_{i_0-1}0)_2 = \frac{\omega}{2^{i_0-1}}(\alpha_1 \cdots \alpha_{i_0-1})_2$$

and

$$\frac{\omega}{2^{i_0}}((\alpha_1 \cdots \alpha_{i_0-1}\alpha_{i_0})_2 + 1) \leqslant \frac{\omega}{2^{i_0-1}} \frac{(\alpha_1 \cdots \alpha_{i_0-1}0)_2 + 2}{2} = \frac{\omega}{2^{i_0-1}}((\alpha_1 \cdots \alpha_{i_0-1})_2 + 1).$$

And so clearly j satisfies (5) for $i = i_0 - 1$.

Similarly, we can see that every $j \in I_\alpha$ satisfies (5) for every $i \in \{1, \ldots, i_0\}$ such that the corresponding bit in α is 1. For any constant $0 < \delta < 1/2$, the proportion of 0-1 sequences of length i_0 that have δi_0 ones is asymptotically $2^{-(1-H(\delta))i_0}$, where $H(\cdot)$ is the entropy function. For any fixed constant $c > 0$, consider any $J = \{i : i_0' \leqslant i \leqslant i_0\}$ with length $i_0 - i_0' + 1 \geqslant cn$ indexing bit positions $\alpha_{i_0'}, \ldots, \alpha_{i_0}$. Then, for a random $\alpha \in \{0,1\}^{i_0}$, with exponentially small exceptional probability $2^{-\Omega(n)}$, there are $\Omega(n)$ bits $\alpha_i = 1$ in those bit positions $i \in J$. Then any $j \in I_\alpha$ gives the corresponding bit $v_{n-i} = 1$. By (6) this is true under the uniform distribution Pr. for j as well. It follows that with exponentially small exceptional probability $2^{-\Omega(n)}$, a uniformly chosen j defines a number $v = \lfloor \frac{2^n}{\omega} j \rfloor$ with a linear number of bits satisfying $v_{n-i} = 1$, for $i \in J$.

Lemma 3. For any fixed constant $c > 0$ and any $J = \{i : i_0' \leqslant i \leqslant i_0\}$ with length $i_0 - i_0' + 1 \geqslant cn$, there exist constants $c', c'' > 0$ such that, picking a random $0 \leqslant j < \omega$ uniformly, we have $v = \lfloor \frac{2^n}{\omega} j \rfloor$,

$$\mathrm{Pr.}\left(|\{i \in J : v_{n-i} = 1\}| \geqslant c'n\right) \geqslant 1 - 2^{-c''n}.$$

We will write it as

$$\mathrm{Pr.}\left(|\{i \in J : v_{n-i} = 1\}| \geqslant \Omega(n)\right) = 1 - 2^{-\Omega(n)},$$

where the hidden constants in $\Omega(n)$ are uniform for all j in the non-exceptional subset.

Now back to (4) for the probability of observing $|v\rangle$ when noise is present. Regardless what values

$$\sum_{t=1}^{n} \frac{\sum_{s=0}^{n-t} u_{n-t-s}^{(k)} v_s}{2^t}, \quad \text{and} \quad r_1^{(0)}, \ldots, r_{n-b}^{(0)}, r_1^{(1)}, \ldots, r_{n-b-1}^{(1)}, \ldots, r_1^{(n-b-1)}$$

[5)] The reason we cut off at i_0 is to avoid having to deal with intervals that are too small and such odd indexed segments may just miss most integers. We can afford to cut off at i_0, and still get a linear number $\Omega(n)$ of 1's in the first i_0 bits of the n bit binary expansion. This is where we use the fact that ω is large.

are, let us consider only those terms

$$\frac{2\pi\epsilon}{2^b}\left(u_{n-b}^{(k)}v_0r_0^{(0)}+u_{n-b-1}^{(k)}v_1r_0^{(1)}+\ldots+u_0^{(k)}v_{n-b}r_0^{(n-b)}\right)=\frac{2\pi\epsilon}{2^b}\sum_{i=b}^{n}u_{i-b}^{(k)}v_{n-i}r_0^{(n-i)}. \tag{7}$$

We will further throw away some noise terms in (7). Let $d=\mathrm{ord}_2(\omega)$. Recall that $d<\frac{\log_2\omega}{2}$ and $i_0=\lfloor\frac{3}{4}\log_2\omega\rfloor$. Thus, assume b is $O(\log n)$, $i_0-b-d=\Omega(n)$, and we will only consider the subsum in (7) for $i\in\{d+b,\ldots,i_0\}$, which has $\Omega(n)$ terms.

By Lemma 3, except for an exponentially small fraction $2^{-\Omega(n)}$ of j indexing $v=\lfloor\frac{2^n}{\omega}j\rfloor$ $(1\leqslant j<\omega)$, each j defines a linear sized $T_j=\{d+b\leqslant i\leqslant i_0: v_{n-i}=1\}$ (of cardinality $>\Omega(n)$, where the hidden constant in $\Omega(n)$ is uniform for the non-exceptional j's) such that $v_{n-i}=1$ and so $u_{i-b}v_{n-i}r_0^{(n-i)}=u_{i-b}r_0^{(n-i)}$, for $i\in T_j$. Thus we will further ignore a large portion of the above sum (7), and consider only

$$\frac{2\pi\epsilon}{2^b}\sum_{i\in T_j}u_{i-b}^{(k)}r_0^{(n-i)}. \tag{8}$$

Intuitively, any term that was omitted but which in fact survives (i.e., with $u_{i-b}^{(k)}v_{n-i}=1$) can only increase the noise. (Formally, when we eventually apply Lemma 2, these will all be part of the term φ_k.)

Our next goal is to show that, among $i\in T_j$, most pairs of $u^{(k)}=u^*+k\omega$ and $u^{(k')}=u^*+k'\omega$, for $k\neq k'$, have a linear number of different bit values $u_{i-b}^{(k)}\neq u_{i-b}^{(k')}$, for $i\in T_j$.

To investigate the (least i_0-b+1 significant) bits $u_0^{(k)},u_1^{(k)},\ldots,u_{i_0-b}^{(k)}$ of $u^{(k)}=u^*+k\omega$, we consider $u^{(k)} \bmod 2^{i_0-b+1}$. If ω is odd, then ($k\omega \bmod 2^{i_0-b+1}$) will enumerate all values in $\{0,1,\ldots,2^{i_0-b+1}-1\}$ exactly once, when $k=0,1,\ldots,2^{i_0-b+1}-1$. Our range of k is actually from 0 to just below $\frac{2^n-u^*}{\omega}\approx 2^n/\omega\gg 2^{i_0}$. Thus, for any u^*, ($u^*+k\omega \bmod 2^{i_0-b+1}$) enumerates every value in $\{0,1,\ldots,2^{i_0-b+1}-1\}$ almost uniformly.

In general, let $d=\mathrm{ord}_2(\omega)$. Recall from the beginning of Section 3 we may assume $0\leqslant d<\frac{\log_2\omega}{2}$ (which is true with high probability by Lemma 1). Thus $i_0-b-d=\Omega(n)$ for $b=O(\log n)$. The least significant d bits of $k\omega$ are all 0. Therefore, for any u^*, the least significant d bits of $u^{(k)}=u^*+k\omega$ are the same as those of u^*, for all k. Taking away all powers of 2 in ω, we have an odd $\omega'=\omega/2^d$, and thus invertible in the multiplicative group $\mathbb{Z}^*_{2^{i_0-b+1-d}}$. The most significant $i_0-b+1-d=\Omega(n)$ bits in ($u^{(k)} \bmod 2^{i_0-b+1}$) are the same as the bits of $(\lfloor\frac{u^*}{2^d}\rfloor+k\omega') \bmod 2^{i_0-b+1-d}$. And so these $i_0-b+1-d$ bits run through every bit sequence of length $i_0-b+1-d$ exactly once if k runs through $2^{i_0-b+1-d}$ consecutive integers. These are the most significant $i_0-b+1-d$ of the i_0-b+1 least significant bits of $u^{(k)}$.

Consider $u^{(k)}=u^*+k\omega$ and $u^{(k')}=u^*+k'\omega=u^{(k)}+(k'-k)\omega$. For any k, let k' run through $\{0,\ldots,\lfloor\frac{2^n-u^*}{\omega}\rfloor\}$, then $k'-k$ runs through $\{-k,\ldots,\lfloor\frac{2^n-u^*}{\omega}\rfloor-k\}$, a set of consecutive integers of size $\geqslant\lfloor\frac{2^n}{\omega}\rfloor\geqslant 2^{i_0-b+1-d}(\frac{2^n}{2^{i_0}\omega})$. As $2^{i_0}\omega\leqslant 2^{7n/8}$ we have $\frac{2^n}{2^{i_0}\omega}\geqslant 2^{n/8}$. Hence, $(k'-k)\omega' \bmod 2^{i_0-b+1-d}$ picks every value in $\{0,\ldots,2^{i_0-b+1-d}-1\}$ with probability $\frac{1}{2^{i_0-b+1-d}}\cdot(1\pm 2^{-\Omega(n)})$ over k' taking a uniform value in $\{0,\ldots,\lfloor\frac{2^n-u^*}{\omega}\rfloor\}$. Thus, the most significant $i_0-b+1-d=\Omega(n)$ of the i_0-b+1 least significant bits of $u^{(k')}-u^{(k)}=(k'-k)\omega$ are almost uniform, as k' runs through $\{0,\ldots,\lfloor\frac{2^n-u^*}{\omega}\rfloor\}$. Then using the same argument with the entropy function $H(\cdot)$, for all except a fraction of $2^{-\Omega(n)}$ of the pairs (k,k'), we have $u_{i-b}^{(k)}\neq u_{i-b}^{(k')}$, for all i belonging to a subset of T_j of cardinality $\geqslant c_0n$, where the constant $c_0>0$ is uniform for (k,k').

Lemma 4. Assume $|T_j|=\Omega(n)$, where the hidden constant in $\Omega(n)$ does not depend on j. There exists $c_0>0$, such that for random pairs (k,k'),

$$\mathrm{Pr.}\left(\left|\{i\in T_j: u_{i-b}^{(k)}\neq u_{i-b}^{(k')}\}\right|\geqslant c_0n\right)=1-2^{-\Omega(n)},$$

where the hidden constant in $\Omega(n)$ does not depend on j.

It follows that, except for a $2^{-\Omega(n)}$ fraction of pairs (k,k'), the sum

$$\sum_{i=b}^{n}(u_{i-b}^{(k)}-u_{i-b}^{(k')})v_{n-i}r_0^{(n-i)}$$

contains a linear number $n' \geqslant c_0 n$ of uncancelled terms $r_0^{(n-i)}$ where $v_{n-i} = 1$ and $u_{i-b}^{(k)} \neq u_{i-b}^{(k')}$. To apply Lemma 2, we require $(\frac{\epsilon}{2^b})^{-1} < (n')^{1/3}$, or equivalently $b + \log 1/\epsilon < \frac{1}{3} \log n'$. This gives us $n' > (\frac{2^b}{\epsilon})^3$. Note that the scaling quantity $\frac{2\pi\epsilon}{2^b}$ in (8) corresponds to $\frac{2\pi\sigma}{m}$ in Lemma 2. We will take the parameter $t = n'/(\frac{2^b}{\epsilon})^2$ in Lemma 2. Then Lemma 2 applies, and $t > (n')^{1/3}$.

To summarize the error estimates: (i) except with probability $2^{-\Omega(n)}$, we have $\omega > N^{1/3}$ and $\mathrm{ord}_2(\omega) < \frac{\log_2 \omega}{2}$ by Lemma 1; (ii) except for a fraction of $2^{-\Omega(n)}$ of j's, all $v = \lfloor \frac{2^n}{\omega} j \rfloor$ have $|T_j| = \Omega(n)$ by Lemma 3; (iii) except for a fraction of $2^{-\Omega(n)}$ of all pairs (k, k')'s, the index sets of the sums (8) (of random variables $r_0^{(n-i)}$) defined as k and k' all have a symmetric difference with cardinality $\geqslant n' = (2^b/\epsilon)^2 t$, with $t = \Omega(n^{1/3})$, by Lemma 4.

Our goal is to estimate the expectation, over the random choice $x \in \mathbb{Z}_N^*$ (that defines the period $\omega = \omega(x)$ of the function $f(k) = x^k \bmod N$) and all random noise variables $r_\cdot^{(\cdot)}$, of the probability of observing some $|v\rangle$ that has the form $v = \lfloor \frac{2^n}{\omega} j \rfloor$ for some $0 \leqslant j < \omega$. This probability for any v is the square norm expression in (4).

Finally, by linearity of expectation, we estimate the sum of the expectations of the square norm sum (4) indexed by all $v = \lfloor \frac{2^n}{\omega} j \rfloor$. Note that the sum $\sum_{k=0}^{K-1}$ is over K complex numbers of the unit norm, and thus has a norm at most K. With probability $\leqslant 2^{-\Omega(n)}$, (i) may be violated and the sum over all $v = \lfloor \frac{2^n}{\omega} j \rfloor$ of (4) can be at most $\frac{\omega}{2^n K} K^2 = O(1)$. Assuming (i) holds, then the sum of the terms (4) indexed by the $\leqslant 2^{-\Omega(n)}$ fraction of exceptional v's regarding (ii) has value at most $(2^{-\Omega(n)}\omega)\frac{1}{2^n K} K^2 = 2^{-\Omega(n)}$. Assuming (i) and (ii) are both not violated, we apply Lemma 2. By (iii), we get an upper bound

$$\frac{\omega}{2^n K}\left(K + 2^{-\Omega(n)} K^2 + K^2 2^{-\Omega(n^{1/3})}\right) = 2^{-\Omega(n^{1/3})}$$

for the sum of the expectations of (4), where the sum is over the $1 - 2^{-\Omega(n)}$ fraction of non-exceptional $v = \lfloor \frac{2^n}{\omega} j \rfloor$ $(0 \leqslant j < \omega)$ regarding (ii).

We conclude that the expectation (over the random choice $x \in \mathbb{Z}_N^*$ and the random noise variables $r_\cdot^{(\cdot)}$) of the probability of observing a member in $\{|v\rangle : v = \lfloor \frac{2^n}{\omega} j \rfloor, 0 \leqslant j < \omega\}$ is exponentially small.

The proof carries over easily to those $|v\rangle$ that are in the vicinity of a polynomial range of $\lfloor \frac{2^n}{\omega} j \rfloor$, for some $0 \leqslant j < \omega$. And since the estimate is exponentially small, the proof shows that the probability of observing any member of the set of those $|v\rangle$ that are polynomially close to any integral multiple of $\frac{2^n}{\omega}$ is still exponentially small in expectation.

4 Pairs of random primes

The proof in Section 3 exhibits a particular set of primes of positive density, and shows that if the input N to Shor's algorithm is of the form $N = pq$ for any primes p and q from that set, then the algorithm does not factor with exponentially small exceptional probability, if the rotational gates are accompanied by a suitable level of noise.

In cryptography, an interesting question concerns the performance on $N = pq$ for random primes p and q of length m. In this section, we prove Theorem 3, dealing with random pairs of primes p and q chosen uniformly from all primes of the same length.

To prove Theorem 3, we will appeal to some number theoretic estimates for the following.

- The period $\omega_N(x)$ of a random element $x \in \mathbb{Z}_N^*$, where $N = pq$, and p and q are primes uniformly randomly chosen from all primes of length m. (The period $\omega_N(x)$ is the order of x as a group element in $\mathbb{Z}_N^*$.)
- The exact order of the prime 2 of the integer $\omega_N(x)$, i.e., $\mathrm{ord}_2(\omega_N(x))$, for a random element $x \in \mathbb{Z}_N^*$, where $N = pq$, and p and q are primes uniformly randomly chosen from all primes of length m.

For primes p and q of binary length m, $N = pq$ has binary length $\approx 2m$, and the QFT circuit uses about $4m$ qubits with $2^{4m} \approx N^2$. The statement $b + \log 1/\epsilon < \frac{1}{3} \log m - c$ for some $c > 0$ is equivalent to $b + \log 1/\epsilon < \frac{1}{3} \log(4m) - c'$ for some $c' > 0$. We note that to carry through the same proof of Theorem 1 for a pair of chosen primes and $x \in \mathbb{Z}_{pq}^*$, we only need to have the property:

(1) $\omega_N(x)$ is large, say $\omega_N(x) = 2^{\Omega(m)}$, and

(2) $\mathrm{ord}_2(\omega_N(x))$ is not too large, say $\mathrm{ord}_2(\omega_N(x)) = o(m)$.

Håstad, Schrift, and Shamir (acknowledging Alon) [27] proved a version of the following theorem (Theorem 5, see Proposition 1 in p.378 in [27]). Their theorem is sufficient to address property (1) for our purpose. But we will give a minor improvement using the Brun-Titchmarsh theorem, which will be used to derive a bound for property (2) as well. The proof will follow essentially the same reasoning as in [27]; the minor improvement comes from using the Brun-Titchmarsh theorem and an estimate due to Rosser and Schoenfeld (Theorem 15 in [28]):

$$\frac{d}{\phi(d)} \leqslant \mathrm{e}^{\gamma} \cdot \log\log d \cdot \left(1 + \frac{2.5}{\mathrm{e}^{\gamma}(\log\log d)^2}\right),$$

where $\phi(\cdot)$ is the Euler totient function, $\gamma = 0.577\cdots$ is Euler's constant, and log denotes natural logarithm (as it will be for the rest of this section). The estimate is valid for every $d \geqslant 3$, except one case $d = 2\times 3\times\cdots\times 23$ when the constant 2.5 should be replaced by 2.50637. We will just use $\frac{d}{\phi(d)} \leqslant C\log\log d$ for some universal constant C, and all $d \geqslant 3$.

Let $X = 2^m - 1$ and $Y = \lceil\frac{X}{2}\rceil = 2^{m-1}$.

Theorem 5. There exists a constant C, such that for any m and any randomly chosen distinct primes p and q of binary length m, $N = p\cdot q$, and let g be a randomly chosen element in $\mathbb{Z}_N^*$, then for all $m^2 \leqslant A < X$,

$$\Pr.\left(\omega_N(g) < \frac{1}{A}\phi(N)\right) \leqslant C\frac{m^{2/5}}{A^{1/5}},$$

where the probability is over primes $p \neq q$ such that $Y \leqslant p, q \leqslant X$, and $g \in \mathbb{Z}_N^*$.

Note that $\phi(N) = (p-1)(q-1) \approx 2^{2m}$. If we take $A = 2^{2\epsilon m}$ then a random $\omega_N(g) \geqslant 2^{2(1-\epsilon)m}$ with probability $1 - O(m2^{-\epsilon m/5})$. This is more than sufficient for our required property (1) above.

The Brun-Titchmarsh theorem is a reasonably sharp estimate for the number of primes up to any upper bound x, in an arithmetic progression. The bound is applicable even when the modulus of the arithmetic progression is large. The following version is an improvement of the original Brun-Titchmarsh theorem proved by Montgomery and Vaughan [29,30]. Suppose a and d are relatively prime. Let $\pi(x; d, a)$ denote the number of primes $p \equiv a \bmod d$, with $p \leqslant x$.

Theorem 6 (Montgomery-Vaughan (Theorem 2 in p.121 of [30])). Let d and a be relatively prime positive integers, and let $x > d$ be any positive integer. Then

$$\pi(x; d, a) \leqslant \frac{2x}{\phi(d)\log(x/d)},$$

where $\phi(\cdot)$ is the Euler totient function, and log denotes natural logarithm.

Following [27], the proof of Theorem 5 is based on two lemmas. Let $O_N = \max\{\omega_N(x) : x \in \mathbb{Z}_N^*\}$ be the exponent of the finite Abelian group $\mathbb{Z}_N^* \cong \mathbb{Z}_p^* \times \mathbb{Z}_q^*$, then $O_N = \mathrm{lcm}(p-1, q-1)$, and $\omega_N(x) | O_N$ for all $x \in \mathbb{Z}_N^*$.

Lemma 5. There exists a constant $C_1 > 0$, such that for randomly chosen distinct primes p and q of binary length m, $N = p\cdot q$, and for any $1 \leqslant A_1 \leqslant X^{1/4} < 2^{m/4}$,

$$\Pr.\left(O_N < \frac{1}{A_1}\phi(N)\right) \leqslant C_1\frac{1}{A_1}.$$

Proof. It is trivial if $m \leqslant 2$. We will assume $m > 2$. Clearly $O_N = \phi(N)/\gcd(p-1, q-1)$. So,

$$O_N < \frac{1}{A_1}\phi(N) \Longleftrightarrow \gcd(p-1, q-1) > A_1.$$

By the Prime Number Theorem, the number of primes of length m is $\pi(X) - \pi(Y) \approx \frac{X}{2\log X}$. And so the number of ordered pairs of distinct primes of length m is approximately $(\frac{X}{2\log X})^2$. Now we bound the cardinality of

$$S = \{(p, q) : Y \leqslant p \neq q \leqslant X, p, q \text{ are primes, and } \gcd(p-1, q-1) > A_1\}.$$

For $p \neq q$ in that range, we claim that $p - 1 \nmid q - 1$. For otherwise $(q-1)/2 \geqslant p - 1$, which implies that $q \geqslant 1 + 2(Y-1) = X$ and hence $q = X$. Then $p - 1 \leqslant (q-1)/2 = Y - 1 \leqslant p - 1$, and so equality holds, and $p = Y = 2^{m-1}$, a contradiction. It follows that $\gcd(p-1, q-1) \leqslant (p-1)/2 < X/2$. So, $\gcd(p-1, q-1) \leqslant 2^{m-1} - 1$.

We have

$$\begin{aligned}|S| &= \sum_{d=\lfloor A_1\rfloor+1}^{2^{m-1}-1} \sum_{(p,q)} \mathbf{1}_{[\gcd(p-1,q-1)=d]}\\ &\leqslant \sum_{d=\lfloor A_1\rfloor+1}^{2^{m-1}-1} (\pi(X;d,1) - \pi(X/2;d,1))^2,\end{aligned}$$

where $\sum_{(p,q)}$ denotes the sum over primes (p,q) in the range $Y \leqslant p \neq q \leqslant X$, and $\mathbf{1}_{[\gcd(p-1,q-1)=d]}$ is the 0-1 indicator function that is 1 iff $\gcd(p-1,q-1) = d$.

Now we separate the sum into two parts, depending on whether $d > \lfloor A_1^2 X^{1/3} \rfloor$. One part is

$$H = \sum_{d=\lfloor A_1^2X^{1/3}\rfloor+1}^{2^{m-1}-1} (\pi(X;d,1) - \pi(X/2;d,1))^2,$$

where we use the trivial bound $\pi(X;d,1) - \pi(X/2;d,1) \leqslant \frac{X}{2d} + 1$. In the range $d < 2^{m-1}$, it is $\leqslant \frac{X}{d}$. It follows that

$$H < X^2 \sum_{d=\lfloor A_1^2X^{1/3}\rfloor+1}^{\infty} \frac{1}{d^2} < \frac{X^2}{A_1^2X^{1/3}} = \frac{X^{5/3}}{A_1^2}$$

by a comparison to the integral $\int_K^\infty \frac{1}{x^2}\mathrm{d}x = \frac{1}{K}$.

The other part is

$$L = \sum_{d=\lfloor A_1\rfloor+1}^{\lfloor A_1^2X^{1/3}\rfloor} (\pi(X;d,1) - \pi(X/2;d,1))^2,$$

where we use Theorem 6, to get

$$L \leqslant \sum_{d=\lfloor A_1\rfloor+1}^{\lfloor A_1^2X^{1/3}\rfloor} \left(\frac{2X}{\phi(d)\log\frac{X}{d}}\right)^2.$$

As $d \leqslant A_1^2X^{1/3} \leqslant X^{5/6}$, we have $\frac{X}{d} \geqslant X^{1/6}$, and $\log\frac{X}{d} \geqslant (\log X)/6$. So

$$L \leqslant 144 \left(\frac{X}{\log X}\right)^2 \sum_{d=\lfloor A_1\rfloor+1}^{\lfloor A_1^2X^{1/3}\rfloor} \frac{1}{\phi(d)^2}.$$

Next we claim the following.

Claim. $\sum_{d>D} \frac{1}{\phi(d)^2} = O(\frac{1}{D})$, for any $D \geqslant 1$.

To prove this Claim we need a result from Eq. (2.32) in p.61 of [31]:

$$\sum_{n\leqslant x} \left(\frac{n}{\phi(n)}\right)^2 = O(x),$$

for all $x > 0$. Let $a_n = \frac{1}{n^2}$, $b_n = (\frac{n}{\phi(n)})^2$, and $B_n = \sum_{k=D+1}^n b_k$, with $n \geqslant D$. Then $B_D = 0$ and $b_n = B_n - B_{n-1}$, for all $n > D$. We have for all $Z > D$,

$$\sum_{n=D+1}^{Z} \frac{1}{\phi(n)^2} = \sum_{n=D+1}^{Z} a_nb_n = a_ZB_Z + \sum_{n=D+1}^{Z-1} (a_n - a_{n+1})B_n.$$

Now $a_ZB_Z = O(1/Z)$, $a_n - a_{n+1} < 2/n^3$, and thus $(a_n - a_{n+1})B_n = O(1/n^2)$. It follows that

$$\sum_{n=D+1}^{Z} \frac{1}{\phi(n)^2} = O(1/Z) + O(1/D).$$

Letting $Z \to \infty$ proves the Claim.

It follows that

$$L = O\left(\left(\frac{X}{\log X}\right)^2 \cdot \frac{1}{A_1}\right).$$

And

$$|S| \leqslant L + H = O\left(\left(\frac{X}{\log X}\right)^2 \cdot \frac{1}{A_1}\right) + \frac{X^{5/3}}{A_1^2}.$$

Hence,

$$\text{Pr.}\left(O_N < \frac{1}{A_1}\phi(N)\right) = O\left(\frac{1}{A_1}\right).$$

Lemma 5 is proven.

Lemma 6. There exists a constant $C_2 > 0$, such that for any $B > 1$,

$$\text{Pr.}\left(\omega_p(g) < \frac{1}{B}\phi(p)\right) \leqslant C_2\left(\frac{m}{B\log B}\right)^{1/2},$$

where the probability is over a random prime $Y \leqslant p \leqslant X$ and a random $g \in \mathbb{Z}_p^*$, and $\omega_p(g)$ is the order of g as a group element in $\mathbb{Z}_p^*$.

Proof. For any prime p, the order of any $g \in \mathbb{Z}_p^*$ divides the order of the group $\phi(p) = p - 1$,

$$\left|\left\{g \in \mathbb{Z}_p^* : \omega_p(g) < \frac{1}{B}\phi(p)\right\}\right| = \sum_{d|p-1,\ d<\phi(p)/B} \phi(d).$$

Letting $F(p) = \sum_{d|p-1,\ d<\phi(p)/B} \phi(d)$ for any prime p, we have

$$\sum_{Y\leqslant p\leqslant X} F(p) = \sum_{d<X/B} \phi(d) \sum_{Y\leqslant p\leqslant X} \mathbf{1}_{[d|p-1]} \leqslant \sum_{d<X/B} \phi(d)\pi(X; d, 1),$$

where $\mathbf{1}_{[d|p-1]}$ is the 0-1 indicator function. Now we apply Theorem 6 and obtain

$$\sum_{Y\leqslant p\leqslant X} F(p) \leqslant \sum_{d<X/B} \frac{2X}{\log(X/d)} \leqslant \frac{2X^2}{B\log B}.$$

It follows that for any $B' > 0$,

$$|\{p : Y \leqslant p \leqslant X,\ p \text{ is a prime, and } F(p) \geqslant X/B'\}| \leqslant \frac{2X^2}{B\log B} \cdot \frac{B'}{X} = \frac{2XB'}{B\log B}.$$

Then, by the Prime Number Theorem,

$$\text{Pr.}\,(F(p) \geqslant X/B') \leqslant O\left(\frac{B'\log X}{B\log B}\right).$$

Conditional on any p such that $Y \leqslant p \leqslant X$ and $F(p) < X/B'$, the probability over $g \in \mathbb{Z}_p^*$ of the event $\omega_p(g) < \frac{1}{B}\phi(p)$, is $\frac{F(p)}{p-1} < \frac{3}{B'}$. Thus, the conditional probability over both p and $g \in \mathbb{Z}_p^*$ given $F(p) < X/B'$ is

$$\text{Pr.}\left[\omega_p(g) < \frac{1}{B}\phi(p) \middle| F(p) < \frac{X}{B'}\right] = O\left(\frac{1}{B'}\right).$$

It follows easily that

$$\text{Pr.}\left(\omega_p(g) < \frac{1}{B}\phi(p)\right) = O\left(\frac{B'\log X}{B\log B}\right) + O\left(\frac{1}{B'}\right).$$

Setting $B' = (B\log B/\log X)^{1/2}$, gives the bound of Lemma 6.

Now the proof of Theorem 5 can be completed.

Proof. (of Theorem 5) We will pick A_1 and A_2 such that $A = A_1A_2$, then

$$\Pr.\left(\omega_N(g) < \frac{1}{A}\phi(N)\right) = \Pr.\left(O_N < \frac{1}{A_1}\phi(N)\right) + \Pr.\left(\omega_N(g) < \frac{1}{A_2}O_N\right),$$

where the first expression is over primes $Y \leqslant p \neq q \leqslant X$ and the second expression is over p, q, and $g \in \mathbb{Z}_N^*$. This is seen by the contrapositive: if $\phi(N) \leqslant A_1 O_N$ and $O_N \leqslant A_2\omega_N(g)$ then $\phi(N) \leqslant A\omega_N(g)$.

By Lemma 5, the first term is $O(\frac{1}{A_1})$.

For the second term, we know that $\omega_N(g) = \mathrm{lcm}(\omega_p(g), \omega_q(g))$, as $\mathbb{Z}_N^* \cong \mathbb{Z}_p^* \times \mathbb{Z}_q^*$. $\omega_p(g)$ is a divisor of $p-1$, and similarly for $\omega_q(g)$. We write $\omega_p(g) = (p-1)/a$, and $\omega_q(g) = (q-1)/b$, then

$$\omega_N(g) \geqslant \frac{\mathrm{lcm}(p-1, q-1)}{ab}.$$

To see this, we take any prime $r \mid \mathrm{lcm}(p-1, q-1)$,

$$\begin{aligned}\mathrm{ord}_r(\omega_N(g)) &= \max\{\mathrm{ord}_r(p-1) - \mathrm{ord}_r(a), \mathrm{ord}_r(q-1) - \mathrm{ord}_r(b)\}\\ &\geqslant \max\{\mathrm{ord}_r(p-1), \mathrm{ord}_r(q-1)\} - \max\{\mathrm{ord}_r(a), \mathrm{ord}_r(b)\}\\ &\geqslant \max\{\mathrm{ord}_r(p-1), \mathrm{ord}_r(q-1)\} - \mathrm{ord}_r(ab)\\ &= \mathrm{ord}_r(\mathrm{lcm}(p-1, q-1)) - \mathrm{ord}_r(ab).\end{aligned}$$

It follows that, after taking $B = \sqrt{A_2}$,

$$\Pr.\left(\omega_N(g) < \frac{1}{A_2}O_N\right) \leqslant \Pr.\left(\omega_p(g) < \frac{p-1}{B}\right) + \Pr.\left(\omega_q(g) < \frac{q-1}{B}\right) = O\left(\frac{m}{B\log B}\right)^{1/2},$$

by Lemma 6. Equalizing the two error bounds we set

$$\frac{1}{A_1} \approx \left(\frac{m}{B\log B}\right)^{1/2},$$

subject to $1 \leqslant A_1 \leqslant X^{1/4}$, $A_1B^2 = A$, $B > 1$, where A is given as $m^2 \leqslant A < X$.

We can set $B = \frac{(A^2m)^{1/5}}{\log A}$ to achieve the bound in Theorem 5.

We remark that, for polynomial bounded $A = m^k$, we can choose B slightly better, $B = (\frac{m^{2k+1}}{\log m})^{1/5}$, and achieve the following.

Theorem 7. With the same setting as in Theorem 5, for any $k \geqslant 2$,

$$\Pr.\left(\omega_N(g) < \frac{1}{m^k}\phi(N)\right) \leqslant O\left(\frac{1}{m^{(k-2)/5}(\log m)^{2/5}}\right),$$

where the probability is over all random $Y \leqslant p \neq q \leqslant X$ and $g \in \mathbb{Z}_N^*$. The constant in O depends on k.

Finally, to finish the proof of Theorem 3, we address the required property (2), again using the Brun-Titchmarsh theorem.

For any prime p, we have the prime factorization $p - 1 = 2^{e_0}p_1^{e_1}\cdots p_k^{e_k}$. We have

$$\Pr.\left(\exists g \in \mathbb{Z}_p^* : \mathrm{ord}_2(\omega_p(g)) \geqslant e\right) \leqslant \frac{1}{\pi(X) - \pi(Y)}\frac{2X}{\phi(2^e)\log(X/2^e)},$$

where the probability is over a random $Y \leqslant p \leqslant X$.

We have $\phi(2^e) = 2^{e-1}$ for $e \geqslant 1$, and $\pi(X) - \pi(Y) = \Theta(X/\log X)$. Then

$$\Pr.\left(\exists g \in \mathbb{Z}_p^* : \mathrm{ord}_2(\omega_p(g)) \geqslant e\right) \leqslant O\left(\frac{\log X}{2^e\log(X/2^e)}\right).$$

If we set $m^c = 2^e$, then we get an upper bound of $O\left(\frac{1}{m^c}\right)$, where the constant in O depends on c. Thus, for any $c > 0$,

$$\Pr.\left(\exists g \in \mathbb{Z}_p^* : \mathrm{ord}_2(\omega_p(g)) \geqslant c\log_2 m\right) \leqslant O\left(\frac{1}{m^c}\right).$$

As $\omega_N(g) = \mathrm{lcm}(\omega_p(g), \omega_q(g))$, it follows

$$\Pr.\left(\exists g \in \mathbb{Z}_N^* : \mathrm{ord}_2(\omega_N(g)) \geqslant c \log_2 m\right) \leqslant O\left(\frac{1}{m^c}\right).$$

Since both required properties (1) and (2) are separately true with a probability approaching 1, they are jointly true with a probability approaching 1.

5 Some comments

This section contains some comments and personal opinions. They are speculative, and are not to be conflated with the provable part.

Quantum mechanics is unquestionably an accurate model of microscopic physical reality. However, I believe every physical theory is an approximate description of the real world, and quantum mechanics is no exception. In particular, I believe the SU(2) description of possible operations of a qubit to be only approximately true. Specifically, I do not believe arbitrarily small angles have physical meaning.

The real numbers $\mathbb{R}$, the continuum, is a human logical construct in terms of Dedekind cut or Cauchy sequence in the language of ϵ-δ definition. SU(2) (or equivalently SO(3)) as a group, is built on top of the continuum. That these mathematical objects provide remarkable fit in some mathematical theory for physical reality, is an extraordinary fact. But this extraordinary fit is always within a certain range; its unlimited extrapolation is mathematical idealization. For example, the agreement within 10^{-8} between experiments and what the theory quantum electrodynamics (QED) predicts for the electromagnetic fine-structure constant α makes QED one of the most accurate physical theories. The Schrödinger equation $i\hbar\frac{\mathrm{d}}{\mathrm{d}t}|\Psi(t)\rangle = \hat{H}|\Psi(t)\rangle$ suggests that small angles are related to small time periods. But physicists have suggested that time ultimately also comes in discrete and indivisible "units". The concept "chronon" has been proposed as a quantum of time [32]. It has even been proposed that one chronon corresponds to about 6.27×10^{-24} s for an electron, much longer than the Planck time, which is only about 5.39×10^{-44} [33] (see also [34, 35]). (Of course the literal form of the mathematical meaning of Schrödinger equation, as a differential equation, suggests time is infinitely divisible. But my personal view is that this is just mathematical abstraction.)

Thus, I view arbitrarily small angles permitted under SU(2) as mere mathematical abstraction. It is true that using a fixed finite set of rotations of reasonable angles such as $\pi/8$ along various axes can compose rotations of arbitrarily small angles. Quantum error-correction code and the threshold theorems [10, 13–18] ultimately assume that SU(2) and its group composition rule are infinitely accurate, and represent it in a high dimensional tensor product space. The higher the accuracy required, the higher the dimension of the representing tensor product space. Since I doubt the mathematical abstraction of SU(2) is infinitely accurate, I also doubt the composition of a sequence of group elements in SU(2) corresponds infinitely accurately to physical reality. There is a further concern that, when a large number of particles are present, whether the tensor product space representation is infinitely accurate. In fact, it is conceivable that the great effort in quantum computing will one day lead to modifications of the theory needed for huge multi-particle systems. Based on these considerations, it seems to me that permitting some noise in the model is reasonable. The random noise model in this paper is just a model, is not meant as reality.

Finally, in the near to intermediate term, there is the reality that one cannot yet achieve meaningful error correction in actual quantum computers. Noisy intermediate scale quantum (NISQ) [36] technology has received a lot of attention. The proof in this paper places a hard limit where failure provably occurs without quantum error correction.

Of course, in addition to its intrinsic interest, factoring integers of the form pq is at the heart of the Rivest-Shamir-Adleman (RSA) public-key cryptosystem [37]. However several results and conjectures in number theory suggest that the failure reported in this paper of Shor's factoring algorithm in the presence of noise can be more severe in the asymptotic sense. We used a theorem of Fouvry [24] to produce a set of primes of positive density that have the desired properties of the period of a random element. The most important property is that this period is sufficiently large. In Theorem 3 we proved a version of the theorem for primes of density one. There are deep results and many conjectures about the distribution of prime factorizations of $p - 1$. In the extreme there are the so-called Sophie Germain primes p' such that $p = 2p' + 1$ is also a prime. It is conjectured that there are $2C\frac{x}{(\log_e x)^2}$ many Sophie Germain primes up to x, where $C = \prod_{p>2} \frac{p(p-2)}{(p-1)^2} \approx 0.660161$ is the Hardy-Littlewood twin prime constant. This is just slightly less than positive density. (However, it has not been proven that there are infinitely many Sophie Germain primes.) Sophie Germain primes were studied in (the first case of) Fermat's Last Theorem. Indeed, Adleman and Heath-Brown [38], and Fouvry [24] proved that the first case of Fermat's Last Theorem holds for infinitely many primes p (see also [39]). Another property we use of primes of the property of Theorem 4 is that the period of a random element in $\mathbb{Z}_N^*$ does not have high ord_2. Erdös and Odlyzko [40] proved that the set of odd divisors of $p-1$ has a positive density.

The core of the analysis of Theorem 1 is to deal with a convolutional sum of bits, in the form of $\sum_i u_i^{(k)} v_{n-i} r^{(n-i)}$ (see (7)), where the bits of $u^{(k)}$ come from an arithmetic progression. We essentially have to show that they behave quite "generic", and that distinct terms $u^{(k)}$ and $u^{(k')}$ of the arithmetic progression behave somewhat independently (all with suitable various exceptions). This accords with our intuition. However, such intuition can be faulty sometimes. For example, Newman [41] showed that for the binary bits of the numbers in multiples of three: $3, 6, 9, \ldots, 3k, \ldots$, there is a definite preponderance of those containing an even number of ones over those containing an odd number of ones. Therefore, for these problems intuitive plausibility is not sufficient; a proof as presented in this paper is needed.

Lastly, we give a few comments on the Strong Church-Turing thesis. It is conceivable that some other quantum algorithms in the BQP model can factor integers (or some other seemingly difficult problems) in polynomial time, and withstand the random noise discussed in this paper. Separately, it is definitely conceivable that at some future time, a quantum algorithm is superior to the best "classical" factoring algorithms for integers of a certain range. However I am not convinced that quantum computing as formulated by BQP requires that we modify the Strong Church-Turing thesis, even if factoring is eventually known to be outside P or BPP. In Turing's careful definition of computability, he made a deliberate choice that the "primitive" steps of such a computing device must be discrete. Thus, the set of states of a Turing machine (TM) is finite; the symbols are placed in discrete cells; the alphabet set is finite. At its most fundamental level, it is not permitted to ask the computing machine to scan and differentiate with infinite precision a continuously deformed symbol, say from ξ to ζ, while a mathematical homotopy can easily be envisioned. I believe the model BQP, in its use of the full SU(2) as primitive steps (or what amounts to, equivalently, the assumption that the exact rule of composition of SU(2) corresponds exactly to realizable computational steps), is a departure from the Turing model.

Acknowledgements This work was finished while the author was attending the International Congress of Basic Science in Beijing. The author wants to express his sincere gratitude to the anonymous referees for their careful reading and comments. He also thanks Al AHO, Dan BONEH, Péter GÁCS, Zvi GALIL, Fred GREEN, Steve HOMER, Gil KALAI, Leonid LEVIN, Dick LIPTON, Ashwin MARAN, Albert MEYER, Ken REGAN, Ron RIVEST, Peter SHOR, Mike SIPSER, Les VALIANT, and Ben YOUNG for insightful comments. And he particularly thanks Eric BACH for inspiring discussions on some of the number theoretic estimates. A similar result can be proven for Shor's algorithm computing discrete logarithm, and will be reported later.

References

1 Shor P W. Algorithms for quantum computation: discrete logarithms and factoring. In: Proceedings of the 35th Annual Symposium on Foundations of Computer Science, 1994. 124–134

2 Shor P W. Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. SIAM J Comput, 1997, 26: 1484–1509

3 Landauer R. Information is physical, but slippery. In: Quantum Computing and Communications. London: Springer, 1999. 59–62

4 Landauer R. Is quantum mechanics useful? In: Ultimate Limits of Fabrication and Measurement. Dordrecht: Springer, 1995

5 Unruh W G. Maintaining coherence in quantum computers. Phys Rev A, 1995, 51: 992–997

6 Aaronson S. Quantum Computing Since Democritus. Cambridge: Cambridge University Press, 2013

7 Shor P. Scheme for reducing decoherence in quantum computer memory. Phys Rev, 1995, 52: 2493–2496

8 Steane A. Multiple particle interference and quantum error correction. Proc Royal Soc London Ser, 1996, 452: 2551–2573

9 Calderbank A R, Rains E M, Shor P M, et al. Quantum error correction via codes over GF(4). IEEE Trans Inform Theor, 1998, 44: 1369–1387

10 Daniel G. An introduction to quantum error correction and fault-tolerant quantum computation. 2009. ArXiv:0904.2557

11 Nielsen M A, Chuang I L. Quantum Computation and Quantum Information. Cambridge: Cambridge University Press, 2010

12 Shor P W. Fault-tolerant quantum computation. In: Proceedings of the 37th Conference on Foundations of Computer Science, 1996. 56–65

13 Aharonov D, Ben-Or M. Fault-tolerant quantum computation with constant error. In: Proceedings of the 29th Annual ACM Symposium on Theory of Computing, 1997. 176–188

14 Kitaev A Y. Quantum computations: algorithms and error correction. Russ Math Surv, 1997, 52: 1191–1249

15 Aliferis P. Level reduction and the quantum threshold theorem. Dissertation for Master's Degree. Pasadena: California Institute of Technology, 2007

16 Steane A M, Lucas D M. Quantum computing with trapped ions, atoms and light. Fortschr Phys, 2000, 48: 839–858

17 Knill E. Quantum computing with realistically noisy devices. Nature, 2005, 434: 39–44

18 Aliferis P, Gottesman D, Preskill J. Quantum accuracy threshold for concatenated distance-3 code. Quantum Inform Comput, 2006, 6: 97–165

19 Nam Y S, Blümel R. Robustness of the quantum Fourier transform with respect to static gate defects. Phys Rev A, 2014, 89: 042337

20 Fowler A G, Hollenberg L C L. Scalability of Shor's algorithm with a limited set of rotation gates. Phys Rev A, 2004, 70: 032329. Erratum: scalability of Shor's algorithm with a limited set of rotation gates [Phys. Rev. A 70, 032329 (2004)]. Phys Rev A, 2007, 75: 029905

21 Nam Y S, Blümel R. Scaling laws for Shor's algorithm with a banded quantum Fourier transform. Phys Rev A, 2013, 87: 032333

22 Nam Y S, Blümel R. Performance scaling of the quantum Fourier transform with defective rotation gates. Quantum Inf Process, 2015, 16: 721–736

23 Coppersmith D. An approximate Fourier transform useful in quantum factoring. 1994. ArXiv:quant-ph/0201067

24 Fouvry É. Théorème de Brun-Titchmarsh; application au théorème de Fermat. Invent Math, 1985, 79: 383–407

25 Menezes A J, van Oorschot P C, Vanstone S A. Handbook of Applied Cryptography. Boca Raton: CRC Press, 1997

26 Papoulis A, Pillai S U. Probability, Random Variables and Stochastic Processes. 4th ed. Boston: McGraw-Hill, 2002

27 Håstad J, Schrift A W, Shamir A. The discrete logarithm modulo a composite hides $0(n)$ bits. J Comput Syst Sci, 1993, 47: 376–404

28 Rosser J B, Schoenfeld L. Approximate formulas for some functions of prime numbers. Illinois J Math, 1962, 6: 64–94

29 Hooley C. Applications of Sieve Methods to the Theory of Numbers. Cambridge: Cambridge University Press, 1976

30 Montgomery H L, Vaughan R C. The large sieve. Mathematika, 1973, 20: 119–134

31 Montgomery H L, Vaughan R C. Multiplicative Number Theory I: Classical Theory. Cambridge: Cambridge University Press, 2006

32 Margenau H. The Nature of Physical Reality. Boston: McGraw-Hill, 1950

33 Caldirola P. The introduction of the chronon in the electron theory and a charged-Lepton mass formula. Lett Nuovo Cimento, 1980, 27: 225–228

34 Farias R A H, Recami E. Introduction of a quantum of time ("chronon"), and its consequences for quantum mechanics. 1997. ArXiv:quant-ph/9706059

35 Albanese C, Lawi S. Time quantization and q-deformations. J Phys A-Math Gen, 2004, 37: 2983–2987

36 Preskill J. Quantum computing in the NISQ era and beyond. Quantum, 2018, 2: 79

37 Rivest R L, Shamir A, Adleman L. A method for obtaining digital signatures and public-key cryptosystems. Commun ACM, 1978, 21: 120–126

38 Adleman L M, Heath-Brown D R. The first case of Fermat's last theorem. Invent Math, 1985, 79: 409–416

39 Lenstra H W, Stevenhagen J P. Class field theory and the first case of Fermat's last theorem. In: Modular Forms and Fermat's Last Theorem. New York: Springer, 1997

40 Erdös P, Odlyzko A M. On the density of odd integers of the form $(p-1)/2^{-n}$ and related questions. J Number Theor, 1979, 11: 257–263

41 Newman D J. On the number of binary digits in a multiple of three. Proc Amer Math Soc, 1969, 21: 719–721