# UNCONDITIONAL CORRECTNESS OF RECENT QUANTUM ALGORITHMS FOR FACTORING AND COMPUTING DISCRETE LOGARITHMS

CÉDRIC PILATTE

ABSTRACT. In 1994, Shor introduced his famous quantum algorithm to factor integers and compute discrete logarithms in polynomial time. In 2023, Regev proposed a multi-dimensional version of Shor's algorithm that requires far fewer quantum gates. His algorithm relies on a number-theoretic conjecture on the elements in $(\mathbb{Z}/N\mathbb{Z})^\times$ that can be written as short products of very small prime numbers. We prove a version of this conjecture using tools from analytic number theory such as zero-density estimates. As a result, we obtain an unconditional proof of correctness of this improved quantum algorithm and of subsequent variants.

## 1. INTRODUCTION

1.1. **Context and quantum computing results.** Public key cryptography has become a crucial element of our global digital communication infrastructure. Notable examples include the Diffie-Hellman key exchange [6] and the RSA (Rivest-Shamir-Adleman) cryptosystem [19], which rely on the difficulty of finding discrete logarithms and factoring large numbers, respectively.

However, in 1994, Peter Shor [20] developed an algorithm capable of efficiently solving these problems using a quantum computer.

**Theorem** (Shor[1])**.** *There is a quantum circuit having $O(n^2 \log n)$ quantum gates and $O(n \log n)$ qubits with the following property. There is a classical randomised polynomial-time algorithm that solves the factoring problem*

$$\textbf{Input} : \textit{a composite integer } N \leqslant 2^n$$
$$\textbf{Output} : \textit{a non-trivial divisor of } N$$

*using $O(1)$ calls to this quantum circuit, and succeeds with probability $\Theta(1)$.*

Shor's original article [20] also includes a similar algorithm to solve the discrete logarithm problem. This advancement prompted the development of post-quantum cryptography, such as lattice-based cryptography, to ensure the security of our communications in the face of potential quantum computing breakthroughs [4].

Recently, Regev [18] devised a multidimensional variant of Shor's factoring algorithm that reduces the size of the quantum circuit, i.e. the number of quantum gates, to $O(n^{3/2} \log n)$. A distinctive feature of Regev's algorithm is that the quantum circuit must be called $O(\sqrt{n})$ times, rather than a constant number of times for Shor's algorithm. This is not considered to be a serious drawback as the various complexity parameters of the quantum circuit are far more relevant metrics in quantum computing.

However, this remarkable work of Regev initially came with two main limitations.

[1]The version of Shor's algorithm stated here incorporates some small improvements from [21] (bounded number of calls) and [10] (fast integer multiplication). The number of qubits in Shor's algorithm can be brought down to $O(n)$, though this typically requires a larger number of quantum gates (see e.g. [8, 12]).

First, the original version of Regev's algorithm requires $O(n^{3/2})$ qubits – many more than Shor's algorithm. The reason for this ultimately lies in the difficulty of performing classical computations on quantum computers in a space-efficient manner, due to the need for any quantum computation to be *reversible*. In a subsequent paper, Ragavan and Vaikuntanathan [17] improved Regev's algorithm to use only $O(n \log n)$ qubits, while maintaining the circuit size at $O(n^{3/2} \log n)$ quantum gates. This work, inspired by ideas of Kalinski [13], essentially matches the space cost of Shor's algorithm.

The second issue, which we address in this paper, is that Regev's algorithm [18] does not have theoretical guarantees, unlike Shor's algorithm. The correctness of Regev's algorithm is based on an *ad hoc* number-theoretic conjecture which we describe below. This unproven assumption cannot be avoided as it lies at the core of Regev's improvement on the circuit size. The variant of Ragavan and Vaikuntanathan [17] also crucially relies on this conjecture.

In this paper, we prove a version of Regev's conjecture. This allows us to unconditionally prove the correctness of (slightly modified versions of) the algorithms of Regev [18] and Ragavan and Vaikuntanathan [17]. More precisely, we obtain the following algorithmic result.

**Theorem 1.1.** *There is a quantum circuit having $O(n^{3/2} \log^3 n)$ quantum gates and $O(n \log^3 n)$ qubits with the following property. There is a classical randomised polynomial-time algorithm that solves the factoring problem*

$$\textbf{Input} : \text{a composite integer } N \leqslant 2^n$$
$$\textbf{Output} : \text{a non-trivial divisor of } N$$

*using $O(\sqrt{n})$ calls to this quantum circuit, and succeeds with probability $\Theta(1)$.*

This unconditionally establishes the results in [17, 18] up to logarithmic factors.

Regev's algorithm was adapted by Ekerå and Gärtner [7] to the discrete logarithm problem. Their paper [7] also uses the space-saving arithmetic of Ragavan and Vaikuntanathan [17] to obtain a quantum circuit with $O(n^{3/2} \log n)$ gates and $O(n \log n)$ qubits for computing discrete logarithms.

Once more, the correctness of the algorithm by Ekerå and Gärtner [7] relies on an unproven hypothesis, which can be viewed as a stronger form of Regev's conjecture. Our methods also apply to this stronger statement (again, with minor technical adjustments). Thus, we get an analogue of Theorem 1.1 for the discrete logarithm problem, i.e. an unconditional proof of the results in [7] up to logarithmic factors.

**Theorem 1.2.** *There is a quantum circuit having $O(n^{3/2} \log^3 n)$ quantum gates and $O(n \log^3 n)$ qubits with the following property. There is a classical randomised polynomial-time algorithm that solves the discrete logarithm problem*

$$\textbf{Input} : \text{an integer } N \leqslant 2^n \text{ and elements } g, y \in (\mathbb{Z}/N\mathbb{Z})^\times \text{ such that } y \in \langle g \rangle$$
$$\textbf{Output} : \text{an integer } x \text{ such that } g^x \equiv y \pmod{N}$$

*using $O(\sqrt{n})$ calls to this quantum circuit, and succeeds with probability $\Theta(1)$.*

Moreover, the slight technical modifications that we need to introduce to make all these quantum algorithms unconditional are compatible with the error-correction results of [7, 17]. Finally, our results extend to further variants of Regev's algorithm, such as computing discrete logarithms or multiplicative orders modulo $N$ for several elements simultaneously (see [7]).

**Remark 1.3.** By being slightly more careful in the space usage of our quantum algorithms, it is possible to reduce the number of qubits to $O(n \log^2 n)$ for Theorems 1.1 and 1.2. We will not show this here in order to reduce the quantum computing prerequisites to a minimum.

1.2. **An overview of Regev's algorithm.** Let us start by recalling the basic idea behind Shor's factoring algorithm [20].

Let $N$ be the integer to be factored, say odd and with at least two distinct prime factors. The first step is to generate a random integer $1 < a < N$, which can be assumed to be coprime to $N$. Shor proved that there is an efficient quantum algorithm to find the multiplicative order $r$ of $a$ modulo $N$. It can be shown by elementary means that $a^{r/2}$ is a non-trivial square root of 1 modulo $N$ (here we mean that $2 \mid r$ and $a^{r/2} \not\equiv \pm 1$) with probability $\geqslant 1/2$. Whenever this is the case, we get a non-trivial divisor of $N$, namely $\gcd(N, a^{r/2} - 1)$.

The reason that Shor's algorithm uses $O(n^2 \log n)$ quantum gates, where $n := \lceil \log_2 N \rceil$, comes from the part of the quantum circuit that performs modular exponentiation. Consider the task of computing a power $a^M \pmod N$ on a classical computer, where $1 < a < N - 1$ and $M \leqslant N$. This computation can be performed efficiently using the well-known square-and-multiply method, which involves $O(\log M)$ multiplications of two integers modulo $N$. Since two $n$-bit integers can be multiplied in time $O(n \log n)$ by [10], the complexity of this modular exponentiation problem is thus $O(n^2 \log n)$.[2] For Shor's algorithm, a similar modular exponentiation needs to be performed quantumly, which requires $O(n^2 \log n)$ quantum gates.

Regev's improvement of Shor's algorithm is made possible by combining two key ideas.

The first idea in Regev's algorithm is to work in *higher dimensional* space and replace the random parameter $a$ by several integers $b_1, \dots, b_d$ (chosen in a specific way, as we explain below). Eventually, the optimal choice of dimension turns out to be $d \asymp \sqrt{\log N}$. Similarly to Shor's algorithm, the problem of factoring $N$ easily reduces to the task of finding a vector $(e_1, \dots, e_d) \in \mathbb{Z}^d$ such that $\prod_{i=1}^d b_i^{e_i}$ is a non-trivial square root of 1 modulo $N$.

Using additional tools such as the LLL lattice reduction algorithm, Regev [18] generalised Shor's quantum algorithm to efficiently find all such vectors $(e_1, \dots, e_d)$ in a ball of radius $N^{O(1/d)}$ centred at the origin.

Regev's algorithm only succeeds if there indeed exists a vector $v = (e_1, \dots, e_d) \in \mathbb{Z}^d$ such that

(i) $\|v\|_2 \leqslant N^{O(1/d)}$, and

(ii) $\prod_{i=1}^d b_i^{e_i}$ is a non-trivial square root of 1 modulo $N$.

If the parameters $b_1, \dots, b_d$ are "sufficiently multiplicatively independent" modulo $N$, one might heuristically expect the existence of a vector $v = (e_1, \dots, e_d)$ satisfying (i) and (ii). This is, roughly speaking, what Regev needs to assume. To state his conjecture properly, it remains to specify how the parameters $b_1, \dots, b_d$ are chosen. This is a crucial point, as the idea of working in dimension $d$ does not, on its own, offer any advantage over Shor's algorithm.

The second key insight of Regev is to choose $b_1, \dots, b_d$ to be *very small* compared to $N$. We mentioned that the most costly part of Shor's quantum circuit lies in the modular exponentiation step. Similarly, the number of gates of Regev's circuit is dominated by the cost of computing an expression of the form

$$\prod_{i=1}^{d} b_i^{M_i} \pmod N \tag{1}$$

in the quantum setting, where the exponents $M_1, \dots, M_d$ are $\leqslant N^{O(1/d)}$. Regev observed that (1) can be computed classically in time $O(n^{3/2} \log n)$ if $|b_i| \leqslant (\log N)^{O(1)}$ for all $i$ (for $d \asymp \sqrt{n} \asymp \sqrt{\log N}$). This can be achieved by cleverly ordering the intermediate multiplications so that most of them

[2]Note that, even if $a$ is a small integer, say $a = 2$, this procedure still takes time $O(n^2 \log n)$, because most multiplications will involve $n$-bit integers.

involve integers with much fewer than $n$ bits. This classical procedure can then be turned into a quantum version that uses $O(n^{3/2}\log n)$ gates.

Shor's algorithm corresponds to the $d = 1$ case of Regev's algorithm where the parameter $b_1$ is an element of $(\mathbb{Z}/N\mathbb{Z})^\times$ chosen uniformly at random. In this case, simple considerations from elementary number theory immediately imply the existence of an integer $e_1$ satisfying (i) and (ii) with probability $\gg 1$. This would hold more generally for $d \geqslant 1$ if $b_1, \dots, b_d$ were independent, uniformly distributed random elements of $(\mathbb{Z}/N\mathbb{Z})^\times$. However, for Regev's algorithm, the $b_i$'s are far from uniformly distributed in $(\mathbb{Z}/N\mathbb{Z})^\times$ as they are constrained to lie in a very small subset of $(\mathbb{Z}/N\mathbb{Z})^\times$.

1.3. **The number-theoretic conjecture behind Regev's algorithm.** Regev's algorithm [18] and its space-efficient variant [17] crucially rely on the possibility of choosing very small integers $b_1, \dots, b_d$ such that $\prod_{i=1}^d b_i^{e_i}$ is a non-trivial square root of 1 modulo $N$ for some $e_1, \dots, e_d$ with $|e_i| \leqslant N^{O(1/d)}$. We remind the reader that $n \asymp \log N$ and $d \asymp \sqrt{n}$.

The least restrictive bound[3] on the $b_i$'s that still allows for a quantum circuit with $\tilde{O}(n^{3/2})$ gates is to have $|b_i| \leqslant \exp(\tilde{O}(d))$, where the $\tilde{O}(\cdot)$ notation possibly hides a factor $(\log n)^{O(1)}$. Moreover, it is natural to set the $b_i$'s to be prime numbers (as in [7, 17, 18]) in order to avoid obvious multiplicative relations between them.

In this paper, we choose $b_1, \dots, b_d$ to be independent random prime numbers less than $d^{10^3 d}$. With these parameters, we can now state a version of Regev's conjecture[4] that follows from our results.

**Corollary 1.4.** *Let $N > 2$ be an integer. Let $d := \lceil \sqrt{\log N} \rceil$ and $X := d^{10^3 d}$.*

*Let $\mathbf{b}_1, \dots, \mathbf{b}_d$ be i.i.d. random variables, each uniformly distributed in the set of primes $\leqslant X$ not dividing $N$.*

*Then, with high probability, every $x \in \langle \mathbf{b}_1, \dots, \mathbf{b}_d \rangle$ can be expressed as*

$$x \equiv \prod_{i=1}^d \mathbf{b}_i^{e_i} \pmod N$$

*for some integers $e_i$ with $|e_i| \leqslant e^{O(d)}$ for all $1 \leqslant i \leqslant d$.*

The analogue of Regev's algorithm for the discrete logarithm problem, proposed by Ekerå and Gärtner [7], relies on a stronger version of Regev's conjecture. It concerns the geometry of a certain lattice $\mathcal{L}$ which encodes all multiplicative dependencies between the $\mathbf{b}_i$'s (modulo $N$). We prove the following version of it.[5]

**Corollary 1.5.** *Let $N > 2$ be an integer. Let $d := \lceil \sqrt{\log N} \rceil$ and $X := d^{10^3 d}$.*

*Let $\mathbf{b}_1, \dots, \mathbf{b}_d$ be i.i.d. random variables, each uniformly distributed in the set of primes $\leqslant X$ not dividing $N$.*

*Let $\mathcal{L}$ be the random lattice defined by*

$$\mathcal{L} := \Big\{ (e_1, \dots, e_d) \in \mathbb{Z}^d \ : \ \prod_{i=1}^d \mathbf{b}_i^{e_i} \equiv 1 \pmod N \Big\}. \tag{2}$$

*Then, with high probability, this lattice $\mathcal{L}$ has a basis consisting of vectors of Euclidean norm $\leqslant e^{O(d)}$.*

[3] The bound stated in Regev's paper [18] is $|b_i| \leqslant (\log N)^{O(1)}$, but this condition can be relaxed somewhat, as observed by Ragavan (private communication).

[4] Compare with [18, Theorem 1.1] or [17, Conjecture 3.1].

[5] Compare with [7, Assumption 1] or [17, Conjecture E.1]

The main technical result of this paper is Theorem 2.18, of which Corollary 1.5 is a special case. Corollary 1.4 is a direct consequence Corollary 1.5.

1.4. **Subgroup obstructions.** Corollary 1.4 shows that, with high probability, every element $x$ in the subgroup generated by $\mathbf{b}_1, \dots, \mathbf{b}_d$ can be written as a *short* product of these $\mathbf{b}_i$ modulo $N$. To obtain the full strength of Regev's assumption [17, Conjecture 3.1], it would be necessary to show that this subgroup $\langle \mathbf{b}_1, \dots, \mathbf{b}_d \rangle$ contains a non-trivial square root of 1 modulo $N$.

Unfortunately, it is not possible to prove this in full generality without considerable advances on the well-known *least quadratic non-residue problem* in number theory. For any prime number $p$, write $n(p)$ for the smallest positive integer $a$ which is a quadratic non-residue (i.e. not a square) modulo $p$. The best known asymptotic upper bound for $n(p)$ is Burgess's classical result [5] that

$$n(p) \ll_\varepsilon p^{\frac{1}{4\sqrt{e}}+\varepsilon}.$$

For simplicity, suppose that the integer $N$ to be factored is a product of two equally-sized primes $p_1, p_2 \equiv 3 \pmod 4$. Recall that, for Regev's algorithm, the parameters $\mathbf{b}_i$ are chosen to be less than $\exp(\tilde{O}(d))$. With current techniques, we cannot rule out the possibility that both $n(p_1)$ and $n(p_2)$ are larger than this threshold. If this is the case, all $\mathbf{b}_i$ will be quadratic residues modulo both $p_1$ and $p_2$, which implies that $\langle \mathbf{b}_1, \dots, \mathbf{b}_d \rangle$ is contained in the subgroup of squares modulo $N$. In particular, since $p_1, p_2 \equiv 3 \pmod 4$, the subgroup $\langle \mathbf{b}_1, \dots, \mathbf{b}_d \rangle$ does not contain any non-trivial square roots of 1 modulo $N$.

This issue can be approached from several perspectives.

(1) Assuming the Generalised Riemann Hypothesis, Ankeny proved the much stronger bound $n(p) \ll (\log p)^2$ [2]. This suggests that quadratic residues may no longer be an obstruction in this case, and indeed, under GRH, it is straightforward to show that $\langle \mathbf{b}_1, \dots, \mathbf{b}_d \rangle$ contains a non-trivial square root of 1 modulo $N$ with high probability. Assuming GRH would also considerably simplify the proof of Theorem 2.18 and remove a logarithmic factor for the gate and qubit costs in Theorems 1.1 and 1.2. However, in this paper we seek fully unconditional results.

(2) Using a grand zero-density estimate [11, Theorem 1 (1.8)] and the Landau-Page theorem [15, Corollary 11.10], it is possible to prove the above result unconditionally for *almost all* $N$, which would already have very interesting algorithmic consequences. More precisely, it can be shown that there exists a set of exceptions $E \subset \mathbb{N}$ satisfying

$$|E \cap [1, x]| \ll \exp\big((\log x)^{1/2+o(1)}\big)$$

for $x \geqslant 1$, such that for every odd $N \notin E$ with at least two prime factors, the subgroup $\langle \mathbf{b}_1, \dots, \mathbf{b}_d \rangle$ (with $\mathbf{b}_i = \mathbf{b}_i(N)$ as defined earlier) contains a non-trivial square root of 1 modulo $N$ with high probability. From this and Corollary 1.4, it would follow that Regev's algorithm finds a non-trivial divisor of $N$ with high probability for almost all $N$.

(3) In this paper, we follow a different approach to obtain a completely unconditional result that applies to *all* $N$, by slightly modifying the algorithm itself. While the key to the efficiency of Regev's algorithm is that the $\mathbf{b}_i$'s are small, one can tolerate a bounded number of large $\mathbf{b}_i$'s.[6] We use this extra flexibility to overcome the subgroup obstructions. The simplest way to proceed is to allow for one of the parameters, say $\mathbf{b}_1$, to be uniformly distributed in $(\mathbb{Z}/N\mathbb{Z})^\times$. As with Shor's algorithm, this ensures that the subgroup $\langle \mathbf{b}_1 \rangle$, and hence also $\langle \mathbf{b}_1, \dots, \mathbf{b}_d \rangle$, contains a non-trivial square root of 1 with probability $\gg 1$.

[6] In fact, this observation was already needed to adapt Regev's algorithm to the discrete logarithm problem [7].

**Remark 1.6.** In his paper [18], Regev proposed to select the parameters $b_1, \dots, b_d$ deterministically, for example by setting $b_i$ to be the $i$-th smallest prime for $1 \leqslant i \leqslant d$. While such a deterministic choice of parameters is likely to work for almost all $N$, this is probably very difficult to prove.

**Remark 1.7.** In this paper, the parameters $\mathbf{b}_i$ are chosen to be random primes (not dividing $N$) under a certain threshold $X$. Another natural choice would have been to let $\mathbf{b}_1, \dots, \mathbf{b}_d$ be i.i.d. random variables uniformly distributed in the set of all integers $\leqslant X$ coprime to $N$. The proofs in this paper would carry over to this framework if an analogue of Proposition 2.9 for short character sums $\sum_{n \leqslant x} \chi(n)$ were available. Assuming the Generalised Riemann Hypothesis, the work of Granville and Soundararajan [9, Theorem 2] gives a bound of the required strength. However, this approach does not seem sufficient to obtain an unconditional result (the bounds in [9] become too weak when $L(s, \chi)$ has zeroes close to the line $\operatorname{Re} s = 1$).

This paper is organised as follows. Our main technical result is Theorem 2.18, which shows that lattices $\mathcal{L}$ similar to that in Corollary 1.5 have a basis of short vectors with high probability. Using simple geometry of numbers (see Section 2.5), we reduce this problem to estimating the number of lattice points in balls of growing radii. Unfortunately, we are unable to obtain a suitable lattice point count for $\mathcal{L}$ directly. We resolve this by considering a different lattice $\mathcal{L}_M$ from the start of the argument (using the lemmas in Section 2.2). In Section 2.3, we expand the lattice point count for $\mathcal{L}_M$ in terms of Dirichlet characters modulo $N$. This produces a main term, which can be estimated precisely, and an error term. The heart of the proof lies in using a zero-density estimate for Dirichlet characters modulo $N$ to bound this error term unconditionally. Finally, we prove our quantum algorithmic applications (Theorems 1.1 and 1.2) in Section 3.

## Acknowledgements

The author is supported by the Oxford Mathematical Institute and a Saven European Scholarship. I would like to thank Jane Street for awarding me one of their Graduate Research Fellowships. I am indebted to my advisors, Ben Green and James Maynard, for their invaluable advice and support. Finally, I wish to thank Seyoon Ragavan for his kind explanations, and in particular for showing me that the upper bound for the parameters in Regev's algorithm could be relaxed considerably.

## 2. Proof of the existence of a short lattice basis

2.1. **Notation.** We write $f \ll g$ or $f = O(g)$ if $|f| \leqslant Cg$ for some absolute constant $C > 0$. If instead, $C$ depends on a parameter $\theta$, we write $f \ll_\theta g$ or $f = O_\theta(g)$. The notation $f \asymp g$ or $f = \Theta(g)$ means that $f \ll g$ and $g \ll f$.

A *character* of a finite abelian group $G$ is a homomorphism $\chi : G \to \mathbb{C}^\times$. The *order* of $\chi$, denoted by $\operatorname{ord}(\chi)$, is the least positive integer $n$ such that $\chi^n$ is the trivial character 1. The group of all characters of $G$ is denoted by $\widehat{G}$. If $g_1, \dots, g_k \in G$, we write $\langle g_1, \dots, g_k \rangle$ for the subgroup of $G$ generated by these elements.

A non-trivial square root of 1 modulo $N$ is an element $a \in (\mathbb{Z}/N\mathbb{Z})^\times$ such that $a^2 \equiv 1 \pmod N$ and $a \not\equiv \pm 1 \pmod N$. Euler's totient function and the prime counting function are denoted by $\varphi$ and $\pi$, respectively. We use the notation $\|\cdot\|_2$ for the Euclidean norm. We write log for the natural logarithm and $\log_2$ for the logarithm in base 2.

Random variables are typically written in bold font, such as $\mathbf{b}_1, \dots, \mathbf{b}_d$. A statement will be said to occur *with high probability* if it holds with probability tending to 1 as $N \to \infty$.

We refer the reader to [16] for a detailed textbook on quantum computing, and [4, Chapter 2] for a brief overview.

2.2. **Restricting to a convenient subgroup of $\widehat{(\mathbb{Z}/N\mathbb{Z})^\times}$.** The goal of this preliminary section is to define a subgroup of the character group of $(\mathbb{Z}/N\mathbb{Z})^\times$ that does not have too many elements of small order. This will be crucial for Section 2.4, to reduce the influence of potential counterexamples to the Generalised Riemann Hypothesis. For example, we need to avoid having many characters $\chi_1, \dots, \chi_k$ modulo $N$ such that $\chi_i^2 = \psi$ for all $i$, where $\psi$ is an exceptional character (in the sense that $L(s, \psi)$ has a zero very close to the line $\operatorname{Re} s = 1$). The definition of the suitable subgroup depends on the precise structure of $(\mathbb{Z}/N\mathbb{Z})^\times$.

**Notation 2.1.** If $G$ is a finite abelian group (written multiplicatively) and $M \geqslant 1$, we write $G^M := \{x^M : x \in G\}$ for the subgroup of $M$th powers in $G$.[7]

**Lemma 2.2.** *Let $G$ be a finite abelian group (written multiplicatively) and $M \geqslant 1$. There is an isomorphism*

$$\iota : \widehat{G^M} \xrightarrow{\sim} \widehat{G}^M$$

*such that, for any $\chi \in \widehat{G^M}$, $\iota(\chi) = \widetilde{\chi}^M$ where $\widetilde{\chi} \in \widehat{G}$ is an arbitrary extension of $\chi$ to $G$.*

*Proof.* Pontryagin duality for finite abelian groups is an equivalence of categories, and therefore an exact functor. In simple terms, this implies that injections turn into surjections when switching to the dual, and vice versa. Hence, the $M$th power map $G \twoheadrightarrow G^M$, $g \mapsto g^M$, induces an injection $\iota : \widehat{G^M} \hookrightarrow \widehat{G}$ defined by

$$\iota(\chi)(g) := \chi\big(g^M\big)$$

for every $\chi \in \widehat{G^M}$ and $g \in G$. Moreover, the restriction map $\widehat{G} \to \widehat{G^M}$ is surjective, using exactness of Pontryagin duality again. Thus, every $\chi \in \widehat{G^M}$ can be extended to a character $\widetilde{\chi}$ on $G$, and

$$\chi\big(g^M\big) = \widetilde{\chi}(g^M) = \widetilde{\chi}^M(g).$$

This implies that the image of $\iota$ is contained in $\widehat{G}^M$. Since $\big|\widehat{G^M}\big| = \big|G^M\big| = \big|\widehat{G}^M\big|$, the lemma follows. □

**Definition 2.3.** Let $N \geqslant 1$. For every $h \geqslant 1$ we define

$$K(h) := \frac{|(\mathbb{Z}/N\mathbb{Z})^\times|}{|((\mathbb{Z}/N\mathbb{Z})^\times)^h|}.$$

In other words, $K(h)$ is the size of the kernel of the $h$th power map in $(\mathbb{Z}/N\mathbb{Z})^\times$.

**Lemma 2.4.** *Let $N > 2$ be an integer and $d := \lceil \sqrt{\log N} \rceil$.*

*For every prime $p$, define $m_p$ to be the largest non-negative integer such that*

$$K(p^{m_p}) \geqslant p^{dm_p/10}.$$

*Let $M_* = M_*(N) := \prod_p p^{m_p}$. Then the following holds.*

1. *We have $M_* \leqslant e^{10d}$.*
2. *For every $h \geqslant 1$, we have $K(M_* h)/K(M_*) \leqslant h^{d/10}$.*

*Proof.* (1) For every prime $p$, since $K(p^{m_p})$ is a power of $p$ dividing $|(\mathbb{Z}/N\mathbb{Z})^\times| = \varphi(N)$, we have

$$\prod_p K(p^{m_p}) \mid \varphi(N).$$

[7] In particular, $G^M$ is *not* the $M$-fold direct product of $G$.

By definition of $m_p$, we know that $K(p^{m_p}) \geqslant p^{dm_p/10}$ for every $p$, and thus

$$M_*{}^{d/10} = \prod_p p^{dm_p/10} \leqslant \prod_p K(p^{m_p}) \leqslant \varphi(N) \leqslant N.$$

Recalling that $d = \lceil \sqrt{\log N} \rceil$, it follows that $M_* \leqslant e^{10d}$.

(2) In any finite abelian group (written multiplicatively) and for any coprime integers $a$ and $b$, there is an isomorphism

$$\ker\bigl(g \mapsto g^a\bigr) \times \ker\bigl(g \mapsto g^b\bigr) \cong \ker\bigl(g \mapsto g^{ab}\bigr).$$

Thus, if $h = \prod_p p^{e_p}$ is the prime factorisation of $h$, we see that $K(M_*) = \prod_p K(p^{m_p})$ and $K(M_*h) = \prod_p K(p^{m_p+e_p})$, where all products are finite. Whenever $e_p \geqslant 1$, we have

$$K(p^{m_p+e_p}) < p^{d(m_p+e_p)/10}, \quad K(p^{m_p}) \geqslant p^{dm_p/10}$$

by definition of $m_p$. Thus $K(p^{m_p+e_p})/K(p^{m_p}) \leqslant p^{de_p/10}$, and this last inequality also holds when $e_p = 0$. We conclude that

$$\frac{K(M_*h)}{K(M_*)} = \prod_p \frac{K(p^{m_p+e_p})}{K(p^{m_p})} \leqslant \prod_p p^{de_p/10} = h^{d/10}.$$

as claimed. □

**Lemma 2.5.** *Let $N > 2$, $d := \lceil \sqrt{\log N} \rceil$, $G := (\mathbb{Z}/N\mathbb{Z})^\times$ and let $M_* \geqslant 1$ be the integer defined in the statement of Lemma 2.4.*

*Let $h \geqslant 1$. Let $\chi_1, \dots, \chi_n \in \widehat{G^{M_*}}$ be distinct characters such that*

$$\chi_1^h = \chi_2^h = \dots = \chi_n^h,$$

*Then $n \leqslant h^{d/10}$.*

*Proof.* Let $f_h : G^{M_*} \to G^{M_*}$ be the $h$th power map, $f(\chi) := \chi^h$. By assumption, the $n$ distinct characters $\chi_1\chi_n^{-1}, \chi_2\chi_n^{-1}, \dots, \chi_n\chi_n^{-1}$ lie in the kernel of $f$. Hence,

$$n \leqslant \bigl|\ker(f)\bigr| = \frac{|G^{M_*}|}{|(G^{M_*})^h|} = \frac{|((\mathbb{Z}/N\mathbb{Z})^\times)^{M_*}|}{|((\mathbb{Z}/N\mathbb{Z})^\times)^{M_*h}|} = \frac{K(M_*h)}{K(M_*)},$$

which is $\leqslant h^{d/10}$ by Lemma 2.4. □

2.3. **Lattice point counting via characters.** To prove Theorem 2.18, we will need to study the following quantities.

**Definition 2.6.** Let $N \geqslant 1$, $G = (\mathbb{Z}/N\mathbb{Z})^\times$ and $\chi \in \widehat{G}$. For all $d \geqslant 1$ and $b_1, \dots, b_d \in G$, we define the quantity

$$F_{\chi,H}(b_1, \dots, b_d) := \prod_{i=1}^d \sum_{|h| \leqslant H} \chi^h(b_i).$$

The following lemma relates these expressions to a lattice point counting problem.

**Lemma 2.7.** *Let $N, M, d, H \geqslant 1$ be integers. Let $b_1, \dots, b_d \in G := (\mathbb{Z}/N\mathbb{Z})^\times$. The number of vectors $(e_1, \dots, e_d) \in \mathbb{Z}^d \cap [-H, H]^d$ such that*

$$\prod_{i=1}^d b_i^{Me_i} \equiv 1 \pmod N$$

*is equal to*

$$\frac{1}{\left|\widehat{G}^M\right|}\sum_{\chi\in\widehat{G}^M} F_{\chi,H}(b_1,\dots,b_d). \tag{3}$$

*Proof.* By orthogonality of characters in the abelian group $G^M$, we have

$$\mathbf{1}_{\prod_i b_i^{Me_i}\equiv 1 \,(\mathrm{mod}\, N)} = \frac{1}{\left|\widehat{G^M}\right|}\sum_{\psi\in\widehat{G^M}} \psi\left(\prod_{i=1}^d b_i^{Me_i}\right) = \frac{1}{\left|\widehat{G^M}\right|}\sum_{\psi\in\widehat{G^M}}\prod_{i=1}^d \psi\big(b_i^{Me_i}\big).$$

Observe that $\psi\big(b_i^{Me_i}\big) = \iota(\psi)(b_i^{e_i})$, where $\iota$ is the isomorphism $\widehat{G^M}\xrightarrow{\sim}\widehat{G}^M$ defined in Lemma 2.2. We may thus rewrite this equality as

$$\mathbf{1}_{\prod_i b_i^{Me_i}\equiv 1 \,(\mathrm{mod}\, N)} = \frac{1}{\left|\widehat{G}^M\right|}\sum_{\chi\in\widehat{G}^M}\prod_{i=1}^d \chi\big(b_i^{e_i}\big).$$

Therefore, the number of vectors $(e_1,\dots,e_d)\in\mathbb{Z}^d\cap[-H,H]^d$ such that $\prod_{i=1}^d b_i^{Me_i}\equiv 1 \pmod N$ is

$$\sum_{\substack{(e_1,\dots,e_d)\in\mathbb{Z}^d\\ \max_i|e_i|\leqslant H}} \mathbf{1}_{\prod_i \mathbf{b}_i^{Me_i}\equiv 1 \,(\mathrm{mod}\, N)} = \frac{1}{\left|\widehat{G}^M\right|}\sum_{\chi\in\widehat{G}^M}\sum_{\substack{(e_1,\dots,e_d)\in\mathbb{Z}^d\\ \max_i|e_i|\leqslant H}}\prod_{i=1}^d \chi^{e_i}(b_i) = \frac{1}{\left|\widehat{G}^M\right|}\sum_{\chi\in\widehat{G}^M} F_{\chi,H}(b_1,\dots,b_d)$$

as claimed. □

It is fairly straightforward to estimate $F_{\chi,H}(b_1,\dots,b_d)$ when $\chi$ is a character of small order in $\widehat{G}$. The contribution of these small-order characters gives the expected main term for (3).

**Lemma 2.8** (Main term)**.** *Let $N, M, d\geqslant 1$ be integers. Let $b_1,\dots,b_d\in G := (\mathbb{Z}/N\mathbb{Z})^\times$. Uniformly for all integers $H\geqslant e^{31d}$, we have*

$$\frac{1}{\left|\widehat{G}^M\right|}\sum_{\chi\in\widehat{G}^M} F_{\chi,H}(\underline{b}) = \left(1+O\left(\frac{e^{31d}}{H}\right)\right)\frac{(2H+1)^d}{\left|\left\langle b_1^M,\dots,b_d^M\right\rangle\right|} + \frac{1}{\left|\widehat{G}^M\right|}\sum_{\substack{\chi\in\widehat{G}^M\\ \mathrm{ord}(\chi)\geqslant e^{10d}}}|F_{\chi,H}(\underline{b})|, \tag{4}$$

*where $F_{\chi,H}(\underline{b})$ is short for $F_{\chi,H}(b_1,\dots,b_d)$.*

*Proof.* Define

$$\langle b_1,\dots,b_d\rangle^\perp := \{\chi\in\widehat{G}\ :\ \forall i\in[d],\ \chi(b_i)=1\}.$$

Observe that

$$\left|\langle b_1,\dots,b_d\rangle^\perp\cap\widehat{G}^M\right| = \left|\left\{\chi\in\widehat{G^M}\ :\ \forall i\in[d],\ \chi\big(b_i^M\big)=1\right\}\right| = \frac{\left|G^M\right|}{\left|\left\langle b_1^M,\dots,b_d^M\right\rangle\right|} \tag{5}$$

where the first equality follows from Lemma 2.2 and the second from the canonical isomorphism $H_1^\perp\cong\widehat{G_1/H_1}$ for any subgroup $H_1$ of a finite abelian group $G_1$.

Clearly, if $\chi\in\langle b_1,\dots,b_d\rangle^\perp\cap\widehat{G}^M$, we have $F_{\chi,H}(\underline{b}) = (2H+1)^d$. Hence, by (5), those characters contribute

$$\frac{1}{\left|\widehat{G}^M\right|}\sum_{\chi\in\langle b_1,\dots,b_d\rangle^\perp\cap\widehat{G}^M} F_{\chi,H}(\underline{b}) = \frac{(2H+1)^d}{\left|\left\langle b_1^M,\dots,b_d^M\right\rangle\right|}$$

to the sum (4).

For any character $\chi \in \widehat{G}^M \setminus \langle b_1, \dots, b_d\rangle^\perp$ of order $\leqslant e^{10d}$, and any $i \in [d]$ such that $\chi(b_i) \neq 1$, we can bound

$$\left| \sum_{|h| \leqslant H} \chi(b_i)^h \right| \leqslant \operatorname{ord}(\chi) \leqslant e^{10d}$$

by the geometric series formula. Thus, defining

$$I_\chi := \{i \in [d] : \chi(b_i) \neq 1\},$$

we have

$$|F_{\chi,H}(\underline{b})| \leqslant e^{10d|I_\chi|}(2H+1)^{d-|I_\chi|}.$$

Consequently,

$$\sum_{\substack{\chi \in \widehat{G}^M \setminus \langle b_1,\dots,b_d\rangle^\perp \\ \operatorname{ord}(\chi) \leqslant e^{10d}}} |F_{\chi,H}(\underline{b})| \ll \sum_{\substack{I \subset [d] \\ I \neq \emptyset}} e^{10d|I|}(2H+1)^{d-|I|} \sum_{\substack{\chi \in \widehat{G}^M \\ \operatorname{ord}(\chi) \leqslant e^{10d} \\ I_\chi = I}} 1. \tag{6}$$

Fix some non-empty set $I \subset [d]$ and complex numbers $(z_i)_{i \in I}$. Let $C_{I,(z_i)}$ be the set of all characters $\chi \in \widehat{G}^M$ such that

$$\chi(b_i) = \begin{cases} z_i & \text{if } i \in I \\ 1 & \text{if } i \in [d] \setminus I. \end{cases}$$

Note that $C_{I,(z_i)}$ is either the empty set, or a coset of $\langle b_1, \dots, b_d\rangle^\perp \cap \widehat{G}^M$ in $\widehat{G}^M$. Moreover, if $\chi$ has order $\leqslant e^{10d}$, this set $C_{I,(z_i)}$ can only be non-empty if all $z_i$ are roots of unity of order $\leqslant e^{10d}$, and there are $\leqslant e^{20d}$ such roots of unity. We conclude that

$$\sum_{\substack{\chi \in \widehat{G}^M \\ \operatorname{ord}(\chi) \leqslant e^{10d} \\ I_\chi = I}} 1 \leqslant \sum_{(z_i)_{i \in I}} \big|C_{I,(z_i)}\big| \leqslant e^{20d|I|} \big|\langle b_1, \dots, b_d\rangle^\perp \cap \widehat{G}^M\big|.$$

Thus, we can bound the right-hand side of (6) by

$$\ll (2H+1)^d \big|\langle b_1, \dots, b_d\rangle^\perp \cap \widehat{G}^M\big| \sum_{\substack{I \subset [d] \\ I \neq \emptyset}} \left(e^{30d}H^{-1}\right)^{|I|} \ll (2H+1)^d \big|\langle b_1, \dots, b_d\rangle^\perp \cap \widehat{G}^M\big| 2^d e^{30d} H^{-1}$$

where we used that $H \geqslant e^{30d}$ in the last step. By (5), this means that

$$\frac{1}{\big|\widehat{G}^M\big|} \sum_{\substack{\chi \in \widehat{G}^M \setminus \langle b_1,\dots,b_d\rangle^\perp \\ \operatorname{ord}(\chi) \leqslant e^{10d}}} |F_{\chi,H}(\underline{b})| \ll e^{31d} H^{-1} \frac{(2H+1)^d}{\big|\langle b_1^M, \dots, b_d^M\rangle\big|},$$

which completes the proof. □

2.4. **Bounding the contribution of large-order characters.** In this section, we bound the error term coming from large-order characters, on average over primes $b_1, \dots, b_d$ in a short interval.

We will need the following ingredients from classical analytic number theory.

**Proposition 2.9** (Character sums over primes)**.** *Let $1/2 \leqslant \alpha \leqslant 1$. Let $q, x \geqslant 2$. Let $\chi$ be a non-principal character modulo $q$ whose Dirichlet $L$-function $L(s,\chi)$ has no zero in the rectangle*

$$\big\{s \in \mathbb{C} \ : \ \alpha < \operatorname{Re} s \leqslant 1, \ |\operatorname{Im} s| \leqslant x^{1-\alpha}\big\}.$$

*Then*

$$\frac{1}{\pi(x)}\sum_{p\leqslant x}\chi(p) \ll x^{-(1-\alpha)}\log^3(qx).$$

Proposition 2.9 is a standard consequence of the explicit formula for $L(s,\chi)$ and is proved in Appendix A. The logarithmic factors can be somewhat improved, but such refinements are irrelevant here.

The Generalised Riemann Hypothesis is the claim that, for every Dirichlet character $\chi$, $L(s,\chi)$ has no zero $\rho$ with $\frac{1}{2} < \operatorname{Re}\rho < 1$. Assuming GRH, Proposition 2.9 thus implies almost square-root cancellation for character sums over primes. Proposition 2.10 will serve as an unconditional substitute for GRH.

**Proposition 2.10** (Zero-density estimate for a fixed modulus)**.** *Uniformly for $\frac{4}{5} \leqslant \alpha \leqslant 1$ and $q, T \geqslant 1$, we have*

$$\sum_{\chi \,(\mathrm{mod}\, q)} \mathcal{N}(\alpha, T, \chi) \ll_\varepsilon (qT)^{(2+\varepsilon)(1-\alpha)}.$$

*where $\mathcal{N}(\alpha,T,\chi)$ denotes the number of zeros (with multiplicity) of $L(s,\chi)$ in the rectangle*

$$\{s\in\mathbb{C} \;:\; \alpha < \operatorname{Re} s \leqslant 1,\ |\operatorname{Im} s| \leqslant T\}.$$

*Proof.* This is [11, Theorem 1 (1.7)]. □

Our goal is to control the total contribution of all large-order characters in (4). We will do so in Proposition 2.14 (restricting to a suitable sugbroup of $\widehat{G}$). We first prove the following $L^2$ bound for a single large-order character.

**Lemma 2.11.** *Let $N > 2$ be an integer and let $G = (\mathbb{Z}/N\mathbb{Z})^\times$. Let $d = \lceil\sqrt{\log N}\rceil$ and $X = d^{10^3 d}$. Let $\mathbf{b}_1, \dots, \mathbf{b}_d$ be i.i.d. random variables, each uniformly distributed in the set of primes less than $X$ not dividing $N$. Let $\chi \in \widehat{G}$ be a character of order $\geqslant e^{10d}$. Then, for every $H \geqslant e^{10d}$,*

$$\mathbb{E}\big[|F_{\chi,H}(\mathbf{b}_1,\dots,\mathbf{b}_d)|^2\big] \ll d^{-10d}H^{2d}.$$

*Proof.* By Proposition 2.10, the number of Dirichlet characters modulo $N$ whose associated $L$-function has a zero in the region

$$\left\{s\in\mathbb{C} \;:\; 1-\frac{1}{10d} < \operatorname{Re} s \leqslant 1,\ |\operatorname{Im} s| \leqslant X\right\}. \tag{7}$$

is

$$\ll (NX)^{(2+1)/(10d)} \ll e^d. \tag{8}$$

If a non-principal character $\psi$ modulo $N$ has no zero in the region (7), then by Proposition 2.9 we have

$$\mathbb{E}[\psi(\mathbf{b}_1)] \ll X^{-1/(10d)}(\log(NX))^3 \ll d^{-100}d^6 \ll d^{-15}. \tag{9}$$

We now expand $\mathbb{E}\big[|F_{\chi,H}(\mathbf{b}_1,\dots,\mathbf{b}_d)|^2\big]$ as

$$\begin{aligned}\mathbb{E}\left[\Big|\prod_{i=1}^{d}\sum_{|h|\leqslant H}\chi^h(\mathbf{b}_i)\Big|^2\right] &= \prod_{i=1}^{d}\mathbb{E}\left[\Big|\sum_{|h|\leqslant H}\chi^h(\mathbf{b}_i)\Big|^2\right]\\ &= \left(\sum_{|h_1|\leqslant H}\sum_{|h_2|\leqslant H}\mathbb{E}\Big[\chi^{h_1-h_2}(\mathbf{b}_1)\Big]\right)^d\\ &\leqslant (2H+1)^d\left(\sum_{|h|\leqslant 2H}\Big|\mathbb{E}\Big[\chi^h(\mathbf{b}_1)\Big]\Big|\right)^d.\end{aligned}$$

We can use (9) to bound the term $\mathbb{E}\big[\chi^h(\mathbf{b}_1)\big]$, unless $\chi^h$ is principal or $L(s,\chi^h)$ has a zero in the region (7). By (8), the number of values of $h$ with $|h|\leqslant 2H$ for which one of these two situations occurs is

$$\ll \left(1+\frac{H}{\mathrm{ord}(\chi)}\right)e^d \ll e^{-10d}e^dH \ll e^{-d}H.$$

By (9), we deduce that, for all but $O(e^{-d}H)$ values of $|h|\leqslant 2H$, the bound $\mathbb{E}\big[\chi^h(\mathbf{b}_1)\big]\ll d^{-15}$ holds. Hence

$$\mathbb{E}\big[|F_{\chi,H}(\mathbf{b}_1,\dots,\mathbf{b}_d)|^2\big]\leqslant (2H+1)^d\Big(O(e^{-d}H)+O(d^{-15}H)\Big)^d \ll d^{-10d}H^{2d}$$

as desired. □

Lemma 2.11 applies to all characters $\chi$ of order $\geqslant e^{10d}$, but gives a relatively weak upper bound. In Lemma 2.13, we will prove that a stronger bound can be obtained if a small set of exceptions is allowed. We begin by describing the exceptional characters in the following lemma.

**Lemma 2.12.** *There is some absolute constant $t_0\geqslant 1$ such that the following holds.*

*Let $N,G,d,X$ and $\mathbf{b}_i$ be as in Lemma 2.11. Let $H\geqslant e^{10d}$.*

*Let $\chi\in\widehat{G}$ be a character of order $\geqslant e^{10d}$ such that*

$$\mathbb{E}\big[|F_{\chi,H}(\mathbf{b}_1,\dots,\mathbf{b}_d)|^2\big] > e^{-2td}H^{2d}$$

*for some $t$ with $t_0\leqslant t\leqslant 2d$.*

*Then there are integers $-2H\leqslant h_1,h_2\leqslant 2H$ with $0<h_2-h_1\leqslant e^{3t}$ such that both $L(s,\chi^{h_1})$ and $L(s,\chi^{h_2})$ have a zero in the region*

$$\left\{s\in\mathbb{C}\ :\ 1-\frac{t}{100d}<\mathrm{Re}\,s\leqslant 1,\ |\mathrm{Im}\,s|\leqslant X\right\}. \tag{10}$$

*Proof.* As in the proof of Lemma 2.11, we have

$$\mathbb{E}\big[|F_{\chi,H}(\mathbf{b}_1,\dots,\mathbf{b}_d)|^2\big] = \mathbb{E}\left[\Big|\prod_{i=1}^{d}\sum_{|h|\leqslant H}\chi^h(\mathbf{b}_i)\Big|^2\right]\leqslant (2H+1)^d\left(\sum_{|h|\leqslant 2H}\Big|\mathbb{E}\Big[\chi^h(\mathbf{b}_1)\Big]\Big|\right)^d. \tag{11}$$

Let $I_1$ be the set of all $-2H\leqslant h\leqslant 2H$ such that $\chi^h$ is principal. Since $\chi$ has order $\geqslant e^{10d}$, we have

$$|I_1|\ll 1+e^{-10d}H\ll e^{-10d}H.$$

Let $I_2$ be the set of all $-2H\leqslant h\leqslant 2H$ such that $L(s,\chi^h)$ has a zero in the region (10). By contradiction, suppose that the conclusion of Lemma 2.12 does not hold. Then any sub-interval of

$[-2H, 2H]$ of length $e^{3t}$ contains at most one element of $I_2$. This implies that

$$|I_2| \ll 1 + e^{-3t}H \ll e^{-3t}H.$$

Moreover, for any integer $h \in [-2H, 2H] \setminus (I_1 \cup I_2)$, the character $\chi^h$ is non-principal and has no zero in the region (10), which by Proposition 2.9 implies that

$$\mathbb{E}\Big[\chi^h(\mathbf{b}_1)\Big] \ll X^{-t/(100d)}(\log(NX))^3 \ll d^{-10t}d^6 \ll e^{-3t}.$$

Therefore, we can bound

$$\sum_{|h|\leqslant 2H}\Big|\mathbb{E}\Big[\chi^h(\mathbf{b}_1)\Big]\Big| \ll |I_1| + |I_2| + e^{-3t}H \ll e^{-10d}H + e^{-3t}H \ll e^{-3t}H.$$

Hence, by (11) we get

$$\mathbb{E}\big[|F_{\chi,H}(\mathbf{b}_1,\dots,\mathbf{b}_d)|^2\big] \leqslant e^{O(d)}e^{-3td}H^{2d},$$

which contradicts the assumption in the statement if $t_0$ is chosen to be sufficiently large. □

We wish to use a zero-density estimate again to show that there are few characters satisfying the conclusion of Lemma 2.12. For this step to work, we need to restrict to the subgroup $\widehat{G}^{M_*}$ of the full character group $\widehat{G}$, where $M_*$ is the integer defined in Lemma 2.4.

**Lemma 2.13.** *Let $t_0 \geqslant 1$ be the constant from Lemma 2.12. Let $N, G, d, X$ and $\mathbf{b}_i$ be as in Lemma 2.11. Let $H \geqslant e^{10d}$ and let $M_* \geqslant 1$ be the integer defined in Lemma 2.4.*

*For every $t \geqslant t_0$,*

$$\Big|\big\{\chi \in \widehat{G}^{M_*} \ : \ \mathrm{ord}(\chi) \geqslant e^{10d},\ \mathbb{E}\big[|F_{\chi,H}(\mathbf{b}_1,\dots,\mathbf{b}_d)|^2\big] > e^{-2td}H^{2d}\big\}\Big| \ll e^{td/2}. \tag{12}$$

*Proof.* If $t \geqslant 2d$, the bound (12) is trivially satisfied as the right-hand side is $\gg N$. We may thus assume that $t_0 \leqslant t \leqslant 2d$, in which case Lemma 2.12 applies.

Let $E_t$ be the set of all characters $\psi$ modulo $N$ such that $L(s,\psi)$ has a zero in the rectangle defined in (10). By Proposition 2.10, this set $E_t$ has size

$$|E_t| \ll (NX)^{(2+1/2)t/(100d)} \ll e^{td/20}. \tag{13}$$

For every $\chi \in \widehat{G}$ of order $\geqslant e^{10d}$, if $\mathbb{E}\big[|F_{\chi,H}(\mathbf{b}_1,\dots,\mathbf{b}_d)|^2\big] > e^{-2td}H^{2d}$ then by Lemma 2.12 we can write

$$\chi^h = \psi_1\overline{\psi_2}$$

for some integer $0 < h < e^{3t}$ and some characters $\psi_1, \psi_2 \in E_t$. Hence, the left-hand side of (12) is

$$\leqslant \sum_{\psi_1,\psi_2\in E_t}\sum_{0<h<e^{3t}}\Big|\big\{\chi \in \widehat{G}^M \ : \ \chi^h = \psi_1\overline{\psi_2}\big\}\Big|. \tag{14}$$

Since $\widehat{G}^{M_*}$ and $\widehat{G^{M_*}}$ are isomorphic, Lemma 2.5 implies that

$$\Big|\big\{\chi \in \widehat{G}^{M_*} \ : \ \chi^h = \psi_1\overline{\psi_2}\big\}\Big| \leqslant h^{d/10} \leqslant e^{3td/10}. \tag{15}$$

Inserting (13) and (15) into (14), we conclude that the left-hand side of (12) is

$$\leqslant |E_t|^2 e^{3t}e^{3td/10} \leqslant e^{3t}e^{4td/10} \ll e^{td/2}$$

as claimed. □

Combining the previous lemmas, we obtain a suitable bound for the sum of second moments of $F_{\chi,H}(\mathbf{b}_1,\dots,\mathbf{b}_d)$ over all large-order characters in $\widehat{G}^{M_*}$.

**Proposition 2.14** (Large-order characters). *Let $N, G, d, X$ and $\mathbf{b}_i$ be as in Lemma 2.11. Let $M_* \geqslant 1$ be the integer defined in Lemma 2.4. For $H \geqslant e^{10d}$, we have*

$$\sum_{\substack{\chi \in \widehat{G}^{M_*} \\ \mathrm{ord}(\chi) \geqslant e^{10d}}} \mathbb{E}\big[|F_{\chi,H}(\mathbf{b}_1, \dots, \mathbf{b}_d)|^2\big]^{1/2} \ll d^{-2d} H^d.$$

*Proof.* By Lemma 2.11, we know that $\mathbb{E}\big[|F_{\chi,H}(\underline{\mathbf{b}})|^2\big]^{1/2} \leqslant C d^{-5d} H^d$ for all characters $\chi \in \widehat{G}$ of order $\geqslant e^{10d}$, where $C > 0$ is an absolute constant (as before, $\underline{\mathbf{b}}$ stands for $\mathbf{b}_1, \dots, \mathbf{b}_d$). Therefore,

$$\sum_{\substack{\chi \in \widehat{G}^{M_*} \\ \mathrm{ord}(\chi) \geqslant e^{10d}}} \mathbb{E}\big[|F_{\chi,H}(\underline{\mathbf{b}})|^2\big]^{1/2} \leqslant \sum_{m=m_0}^{+\infty} \sum_{\substack{\chi \in \widehat{G}^{M_*} \\ \mathrm{ord}(\chi) \geqslant e^{10d}}} e^{-m+1} H^d \mathbf{1}_{e^{-m} < \mathbb{E}\left[|F_{\chi,H}(\underline{\mathbf{b}})|^2\right]^{1/2} H^{-d} \leqslant e^{-m+1}}(\chi)$$

where $m_0 := \lfloor 5d \log d - \log C + 1 \rfloor$. We may assume that $m_0 \geqslant \max(t_0 d, 4d \log d)$, since otherwise $d \ll 1$ and Proposition 2.14 is trivially satisfied (given that $|\widehat{G}^{M_*}| \leqslant N \ll 1$ when $d \ll 1$).

Applying Lemma 2.13 with $t := m/d$, we obtain that for every $m \geqslant t_0 d$,

$$\Big|\Big\{\chi \in \widehat{G}^{M_*} \ : \ \mathrm{ord}(\chi) \geqslant e^{10d},\ \mathbb{E}\big[|F_{\chi,H}(\underline{\mathbf{b}})|^2\big]^{1/2} > e^{-m} H^d\Big\}\Big| \ll e^{m/2}.$$

Therefore,

$$\sum_{\substack{\chi \in \widehat{G}^{M_*} \\ \mathrm{ord}(\chi) \geqslant e^{10d}}} \mathbb{E}\big[|F_{\chi,H}(\underline{\mathbf{b}})|^2\big]^{1/2} \ll \sum_{m=m_0}^{+\infty} e^{m/2} e^{-m+1} H^d \ll e^{-m_0/2} H^d$$

which is $\ll d^{-2d} H^d$ as $m_0 \geqslant 4d \log d$. □

2.5. **Geometry of numbers.** In this section, we show how to pass from good estimates on the number of lattice points in certain regions to the existence of a short basis for the lattice.

**Lemma 2.15.** *Let $L \geqslant 1$ be an integer. Cover the cube $[-L, L]^d$ by $(2L)^d$ cubes of side length 1 in the obvious way. Label these unit cubes $C_1, \dots, C_{(2L)^d}$ (in any order). Let $V \subset \mathbb{R}^d$ be a hyperplane through the origin. Then the number of unit cubes $C_i$ intersecting $V$ is*

$$\leqslant (d+1)(2L)^{d-1}.$$

*Proof.* Let $e_1, \dots, e_d$ be the standard basis for $\mathbb{R}^d$. Let $v \in \mathbb{R}^d$ be a unit vector orthogonal to $V$. Without loss of generality, since $\|v\|_2 = 1$, we may assume that $\langle v, e_1 \rangle \geqslant 1/\sqrt{d}$.

Suppose that $V$ intersects two unit cubes $C_i$ and $C_j$ where $C_i = C_j + k e_1$ for some integer $k \geqslant 0$. Then, there is some $p \in C_i$ and some $w \in \mathbb{R}^d$ with $\|w\|_\infty \leqslant 1$ such that $p \in V$ and $p + k e_1 + w \in V$. Therefore, $k e_1 + w \in V$ and thus $\langle k e_1 + w, v \rangle = 0$. Noting that

$$\langle k e_1 + w, v \rangle \geqslant k \langle e_1, v \rangle - \|w\|_2 \|v\|_2 \geqslant \frac{k}{\sqrt{d}} - \sqrt{d},$$

we deduce that $k \leqslant d$.

We have thus proved that, for any cube $C_i$, there are at most $d+1$ cubes of the form $C_i + k e_1$ for some $k \in \mathbb{Z}$ which intersect $V$. The lemma follows. □

The next lemma allows us to convert information about the number of lattice points in cubes into the existence of short linearly independent lattice vectors.

**Lemma 2.16.** *Let $d \geqslant 1$. Let $\Lambda \subset \mathbb{R}^d$ be a full-rank lattice. Let $2 \leqslant H_0 < H_1$ be real numbers such that $H_1/H_0$ is an integer. Suppose that, for $i \in \{0,1\}$,*

$$\left|\Lambda \cap [-H_i, H_i]^d\right| = \theta_i \frac{(2H_i+1)^d}{\mathrm{vol}(\mathbb{R}^d/\Lambda)} \tag{16}$$

*where $\theta_0, \theta_1 > 0$ satisfy $\frac{H_1}{H_0} > \frac{\theta_0}{\theta_1} d\left(\frac{5}{2}\right)^d$. Then $\Lambda \cap [-H_1, H_1]^d$ contains $d$ linearly independent vectors.*

*Proof.* By contradiction, suppose that there exists a linear hyperplane $V$ containing all the points $v \in \Lambda \cap [-H_1, H_1]^d$.

Let $L = H_1/H_0$. The large cube $[-H_1, H_1]^d$ can be covered by $(2L)^d$ axis-parallel cubes of side length $H_0$ in the natural way. By Lemma 2.15, we can bound

$$\left|\Lambda \cap [-H_1, H_1]^d\right| \leqslant (d+1)(2L)^{d-1} \sup_C |\Lambda \cap C|,$$

where the supremum runs over all (not necessarily centred) axis-parallel cubes $C \subset \mathbb{R}^d$ of side length $H_0$. If $C$ is such a cube and $v \in \Lambda \cap C$, then $C \subset v + [-H_0, H_0]^d$, which implies that

$$|\Lambda \cap C| \leqslant \left|\Lambda \cap \left(v + [-H_0, H_0]^d\right)\right| = \left|\Lambda \cap [-H_0, H_0]^d\right|.$$

Thus,

$$\left|\Lambda \cap [-H_1, H_1]^d\right| \leqslant (d+1)(2L)^{d-1} \left|\Lambda \cap [-H_0, H_0]^d\right|.$$

Plugging in our lattice point estimate (16), we get

$$\theta_1 (2H_1+1)^d \leqslant (d+1)(2L)^{d-1} \theta_0 (2H_0+1)^d.$$

Using $2H_1 + 1 \geqslant 2LH_0$ and $d+1 \leqslant 2d$, this implies that $L \leqslant \frac{\theta_0}{\theta_1} d\left(2 + \frac{1}{H_0}\right)^d$, contradicting the inequality in the statement of the lemma. □

The linearly independent vectors given by Lemma 2.16 can be upgraded to a genuine basis for $\Lambda$ by standard geometry of numbers, namely Mahler's theorem. We state this fact in a slightly more general situation.

**Lemma 2.17.** *Let $d \geqslant 1$. Let $\Lambda_1, \Lambda_2 \subset \mathbb{R}^d$ be full-rank lattices such that $M\Lambda_1 \subset \Lambda_2$ for some integer $M \geqslant 1$. Suppose that $\Lambda_1$ contains $d$ linearly independent vectors in $[-H, H]^d$ for some $H > 0$. Then, $\Lambda_2$ admits a basis where each basis vector has Euclidean norm $\leqslant d^{3/2} MH$.*

*Proof.* By assumption, there are linearly independent vectors $v_1, \ldots, v_d \in \Lambda_1 \cap [-H, H]^d$. Then $Mv_1, \ldots, Mv_d$ are linearly independent vectors of $\Lambda_2$ such that $\max_{i \in [d]} \|Mv_i\|_2 \leqslant \sqrt{d} MH$.

Let $B \subset \mathbb{R}^d$ be the unit ball for the Euclidean norm. Let $0 < \lambda_1 \leqslant \cdots \leqslant \lambda_d$ be the successive minima[8] of $B$ with respect to $\Lambda_2$. Since $Mv_1, \ldots, Mv_d \in \Lambda_2$ are linearly independent, we deduce from the above that $\lambda_d \leqslant \sqrt{d} MH$.

By Mahler's theorem (see [22, Theorem 3.34]), we conclude that $\Lambda_2$ admits a basis of vectors of Euclidean norm

$$\leqslant d\lambda_d \leqslant d^{3/2} MH,$$

which is what we needed to show. □

[8] See [22, Definition 3.29] for the definition of successive minima.

2.6. **Short basis vectors.** We can now prove our main technical result, which may be of independent interest.

**Theorem 2.18.** *Let $N > 2$ be an integer. Let $d := \lceil \sqrt{\log N} \rceil$ and $X := d^{10^3 d}$.*

*Let $\mathbf{b}_1, \dots, \mathbf{b}_d$ be i.i.d. random variables, each uniformly distributed in the set of primes $\leqslant X$ not dividing $N$. Let $r \geqslant 0$ and let $\mathbf{x}_1, \dots, \mathbf{x}_r$ be arbitrary*[9] *random variables taking values in $(\mathbb{Z}/N\mathbb{Z})^\times$. Then, with probability $1 - O\big(d^{-d}\big)$, the lattice*

$$\mathcal{L} := \Big\{ (e_1, \dots, e_d, f_1, \dots, f_r) \in \mathbb{Z}^{d+r} \ : \ \prod_{i=1}^{d} \mathbf{b}_i^{e_i} \prod_{i=1}^{r} \mathbf{x}_i^{f_i} \equiv 1 \pmod N \Big\}$$

*has a basis consisting of vectors of Euclidean norm $\ll e^{42(d+r)}$.*

**Remark 2.19.** The constant 42 in the exponent is by no means the limit our techniques. For example, using smooth cutoffs in Definition 2.6 would significantly reduce the error term in Lemma 2.8 and hence lower this constant. We have not performed such optimisations to keep the paper as simple as possible.

*Proof of Theorem 2.18.* Let $M := M_*(N)$ be the integer defined in Lemma 2.4. We introduce the auxiliary lattice

$$\mathcal{L}_M := \Big\{ (e_1, \dots, e_d, f_1, \dots, f_r) \in \mathbb{Z}^{d+r} \ : \ \prod_{i=1}^{d} \mathbf{b}_i^{Me_i} \prod_{i=1}^{r} \mathbf{x}_i^{Mf_i} \equiv 1 \pmod N \Big\}.$$

By Lemmas 2.7 and 2.8, we have the lattice point estimate

$$\Big| \mathcal{L}_M \cap [-H, H]^{d+r} \Big| = \frac{\big(1 + O\big(e^{31(d+r)} H^{-1}\big)\big)(2H+1)^{d+r}}{\big| \big\langle \mathbf{b}_1^M, \dots, \mathbf{b}_d^M, \mathbf{x}_1^M, \dots, \mathbf{x}_r^M \big\rangle \big|} + \frac{1}{\big|\widehat{G}^M\big|} \sum_{\substack{\chi \in \widehat{G}^M \\ \mathrm{ord}(\chi) \geqslant e^{10d}}} |F_{\chi,H}(\underline{\mathbf{b}}, \underline{\mathbf{x}})| \tag{17}$$

for every integer $H \geqslant e^{31(d+r)}$, where

$$F_{\chi,H}(\underline{\mathbf{b}}, \underline{\mathbf{x}}) := F_{\chi,H}(\mathbf{b}_1, \dots, \mathbf{b}_d, \mathbf{x}_1, \dots, \mathbf{x}_r) = \Bigg( \prod_{i=1}^{d} \sum_{|h| \leqslant H} \chi^h(\mathbf{b}_i) \Bigg) \Bigg( \prod_{i=1}^{r} \sum_{|h| \leqslant H} \chi^h(\mathbf{x}_i) \Bigg).$$

Since the $\mathbf{x}_i$ have unknown distributions, we will use the trivial bound

$$|F_{\chi,H}(\underline{\mathbf{b}}, \underline{\mathbf{x}})| \leqslant (2H+1)^r |F_{\chi,H}(\mathbf{b}_1, \dots, \mathbf{b}_d)|.$$

Applying Proposition 2.14, we deduce that, for all $H \geqslant e^{10d}$,

$$\sum_{\substack{\chi \in \widehat{G}^{M_*} \\ \mathrm{ord}(\chi) \geqslant e^{10d}}} \mathbb{E}\big[ |F_{\chi,H}(\underline{\mathbf{b}}, \underline{\mathbf{x}})|^2 \big]^{1/2} \ll d^{-2d} (2H+1)^{d+r}.$$

By the $L^2$ triangle inequality and Chebyshev's inequality, this implies that, for fixed $H \geqslant e^{10d}$,

$$\mathbb{P}\Bigg( \sum_{\substack{\chi \in \widehat{G}^M \\ \mathrm{ord}(\chi) \geqslant e^{10d}}} |F_{\chi,H}(\underline{\mathbf{b}}, \underline{\mathbf{x}})| > d^{-d} (2H+1)^{d+r} \Bigg) \ll d^{-2d}. \tag{18}$$

[9] In particular, the $\mathbf{x}_i$ are not assumed to be independent or identically distributed.

Let $H_0 := d\lceil e^{31(d+r)}\rceil$ and $H_1 := \lceil d(d+r)(5/2)^{d+r}\rceil H_0$. We apply (17) and (18) twice, once with $H = H_0$ and once with $H = H_1$ to obtain the following: with probability $1 - O\big(d^{-d}\big)$, the two estimates

$$\Big|\mathcal{L}_M \cap [-H_0, H_0]^{d+r}\Big| = (1 + O(1/d))\frac{(2H_0+1)^{d+r}}{\big|\langle \mathbf{b}_1^M, \dots, \mathbf{b}_d^M, \mathbf{x}_1^M, \dots, \mathbf{x}_r^M\rangle\big|}$$

and

$$\Big|\mathcal{L}_M \cap [-H_1, H_1]^{d+r}\Big| = (1 + O(1/d))\frac{(2H_1+1)^{d+r}}{\big|\langle \mathbf{b}_1^M, \dots, \mathbf{b}_d^M, \mathbf{x}_1^M, \dots, \mathbf{x}_r^M\rangle\big|}$$

simultaneously hold.

We now apply Lemma 2.16. By our choice of $H_0$ and $H_1$, the inequality in the statement is satisfied provided that $d$ is sufficiently large. We can assume that $d$ is large enough as, for $d \ll 1$, we have $N \ll 1$ and Theorem 2.18 is trivially true. Thus, Lemma 2.16 implies that $\mathcal{L}_M \cap [-H_1, H_1]^{d+r}$ contains $d + r$ linearly independent vectors, with probability $1 - O\big(d^{-d}\big)$. By Lemma 2.17, since $M\mathcal{L} \subset \mathcal{L}_M$, we conclude that, with probability $1 - O\big(d^{-d}\big)$, $\mathcal{L}$ admits a basis of vectors of Euclidean norm

$$\ll d^{3/2} M H_1 \ll d^{3/2} e^{10d} d^2 (d+r)(5/2)^{d+r} e^{31(d+r)} \ll e^{42(d+r)}$$

using the bound $M \leqslant e^{10d}$ given by Lemma 2.4. This completes the proof. □

## 3. Applications to quantum computing

In this section, we prove the correctness of efficient quantum algorithms for factoring and for the discrete logarithm problem by applying our version of Regev's number-theoretic conjecture, Theorem 2.18.

### 3.1. **Preparatory lemmas.**

**Lemma 3.1.** *Let $N, d, m \geqslant 2$ be integers. There is a classical algorithm that, given integers $0 \leqslant a_1, \dots, a_d \leqslant 2^m$ and exponents $t_1, \dots, t_d \in \{0, 1\}$, computes the product*

$$\prod_{i=1}^{d} a_i^{t_i} \pmod N$$

*in time $O\big(md(\log d)\log(md)\big)$.*

*Proof.* This is similar to [18, p5] or [17, Lemma 5.6] but for a general value of $m$. Without loss of generality, assume that $d$ is a power of 2, say $d = 2^l$. We proceed to compute this product in a binary tree fashion. Let $T(k)$ be the complexity of multiplying any $k$ of these numbers $a_i^{t_i}$ modulo $N$. Note that $T(2k) \leqslant 2T(k) + O(M(mk))$ where $M(x)$ is the time needed to multiply two integers having at most $x$ bits. By the work of Harvey and van der Hoeven [10], it is known that $M(x) = O(x \log x)$, which leads to the bound

$$T(d) \ll \sum_{j=0}^{l} 2^j M(md/2^{j+1}) \ll \sum_{j=0}^{l} md \log(md/2^{j+1}) \ll md(\log d)\log(md).$$

as claimed. □

We can turn this into a quantum circuit with the following well-known fact.

**Lemma 3.2.** *Any classical circuit can be "compiled" into a reversible quantum circuit that carries out the same computations, with the number of gates and qubits used being proportional to the size of the classical circuit.*

*Proof.* This well-known fact is explained in [17, Section A.1]. □

**Lemma 3.3.** *Let $N$ be a sufficiently large integer. Let $d := \lceil \sqrt{\log N} \rceil$ and $X = d^{10^3 d}$. Let $k = d^4$. Let $\mathbf{n}_1, \dots, \mathbf{n}_k$ be i.i.d. random random variables uniformly distributed in $\{1, \dots, X\}$. Then, the probability that at least $d$ of these $\mathbf{n}_i$ are prime numbers not dividing $N$ is $\geqslant 1 - O(1/N)$.*

*Proof.* Since $N$ has $\ll \log N$ prime factors, and since the number of primes $\leqslant X$ is $\gg X/\log X$, we have

$$\mathbb{P}(\mathbf{n}_1 \text{ is prime and } \mathbf{n}_1 \nmid N) \geqslant \frac{c}{\log X}$$

for some absolute constant $c > 0$. Thus, if $E_i$ is the event that $\mathbf{n}_i$ is not prime or divides $N$, then by the union bound and independence,

$$\mathbb{P}\Big( \bigcup_{\substack{I \subset [k] \\ |I| > k-d}} \bigcap_{i \in I} E_i \Big) \leqslant \sum_{l<d} \binom{k}{l} \Big(1 - \frac{c}{\log X}\Big)^{k-l} \leqslant dk^d e^{-c(k-d)/\log X} \ll e^{-d^2}$$

as needed. □

3.2. **The discrete logarithm problem.** We now prove Theorem 1.2. For convenience, we restate it here.

**Theorem 1.2.** *There is a quantum circuit having $O(n^{3/2} \log^3 n)$ quantum gates and $O(n \log^3 n)$ qubits with the following property. There is a classical randomised polynomial-time algorithm that solves the discrete logarithm problem*

> **Input** : *an integer $N \leqslant 2^n$ and elements $g, y \in (\mathbb{Z}/N\mathbb{Z})^\times$ such that $y \in \langle g \rangle$*
> **Output** : *an integer $x$ such that $g^x \equiv y \pmod N$*

*using $O(\sqrt{n})$ calls to this quantum circuit, and succeeds with probability $\Theta(1)$.*

*Proof of Theorem 1.2.* Suppose we are given an integer $N > 2$ and elements $g, y \in (\mathbb{Z}/N\mathbb{Z})^\times$ such that $y$ lies in the subgroup generated by $g$ in $(\mathbb{Z}/N\mathbb{Z})^\times$. Let $d := \lceil \sqrt{\log N} \rceil$ and $X := d^{10^3 d}$. Let $\mathbf{b}_1, \dots, \mathbf{b}_d$ be i.i.d. random variables, each uniformly distributed in the set of primes $\leqslant X$ not dividing $N$.

Consider the random lattice

$$\mathcal{L}_{\underline{\mathbf{b}},g,y} := \Big\{ (e_1, \dots, e_d, f_1, f_2) \in \mathbb{Z}^{d+2} \ : \ \Big( \prod_{i=1}^d \mathbf{b}_i^{e_i} \Big) g^{f_1} y^{f_2} \equiv 1 \pmod N \Big\}.$$

By Theorem 2.18, with probability $1 - O(d^{-d})$, this lattice has a basis consisting of vectors of Euclidean norm $\ll e^{42d}$.

Computing the discrete logarithm of $y$ with respect to the base $g$ reduces (with a polynomial-time classical algorithm) to computing a short basis for $\mathcal{L}_{\underline{\mathbf{b}},g,y}$. To see why this is the case, suppose that we managed to compute a basis $v_1, \dots, v_{d+2}$ for $\mathcal{L}_{\underline{\mathbf{b}},g,y}$ with $\max_i \|v_i\|_2 \ll e^{42d}$. It is then easy to find a vector in $\mathcal{L}_{\underline{\mathbf{b}},g,y}$ of the form $(0, \dots, 0, x, 1)$ for some integer $x$ (note that such a vector exists since we assume that $y \in \langle g \rangle$). Indeed, this amounts to solving a linear system with integer coefficients. The complexity of solving an integer linear system is polynomial in the dimensions of the matrix and the number of bits of the coefficients, which are both $O(d)$ since $\max_i \|v_i\|_2 \ll e^{42d}$ (see [3]). This yields an integer $x$ such that $g^x \equiv y \pmod N$.

The algorithm for solving the discrete logarithm problem thus proceeds as follows.

(1) Generate $d$ primes $b_1, \ldots, b_d$ independently and uniformly at random in the set of primes $\leqslant X$ not dividing $N$. To do this, it suffices to generate $d^4$ independent random integers $\leqslant X$. By Lemma 3.3, the required conditions will be satisfied for at least $d$ of those with probability $1 - O(1/N)$. This first step takes polynomial time on a classical computer using the AKS primality test [1].

(2) Use the procedure described by Ekerå and Gärtner [7] to obtain a basis for the lattice $\mathcal{L}_{\underline{b},g,y}$. This involves making $O(d)$ calls to a quantum circuit, followed by polynomial-time classical post-processing. This step is now guaranteed to succeed with probability $\Theta(1)$, using the fact that $\mathcal{L}_{\underline{b},g,y}$ has a basis of vectors of Euclidean norm $\ll e^{42d}$ with probability $1 - O(d^{-d})$ (Theorem 2.18).

(3) Compute the discrete logarithm of $y$ in classical polynomial time using this basis, as explained above.

It remains to analyse the gate and space costs of the quantum part of this algorithm.

The only modification needed to the analysis of the quantum circuit in [7] comes from the fact that $b_1, \ldots, b_d$ are not quite as small as in [7]. In [7] (and more generally in [17, 18]), the primes $b_i$ are assumed to have $O(\log d)$ bits. In the present situation, we instead have the bound $b_i \leqslant X = d^{10^3 d}$, i.e. each $b_i$ has $O(d \log d)$ bits.

The only place in [7] where the assumption on the size of the $b_i$'s are used is in [7, Lemma 3]. In turn, the only point in the proof of [7, Lemma 3] where this assumption is needed in when [17, Lemma 4.1] is invoked (as a black box). The proof of [17, Lemma 4.1] is given in [17, Section 5], and the size assumption on the $b_i$'s only comes up in [17, Lemma 5.6].

The result [17, Lemma 5.6] essentially[10] states that there is a quantum circuit using $O(d \log^3 d)$ gates and $O(d \log^3 d)$ qubits to perform the computation of $\prod_{i=1}^{d} a_i^{t_i}$ where $t_i \in \{0, 1\}$ and $a_i$ are integers on $O(\log d)$ bits. This is a special case of Lemma 3.1 (used together with Lemma 3.2) applied with $m = O(\log d)$. In our case, we just apply Lemma 3.1 with $m = O(d \log d)$, together with Lemma 3.2 to convert the classical circuit into a quantum one. This yields a quantum circuit having $O(d^2 \log^3 d)$ gates and $O(d^2 \log^3 d)$ qubits to compute $\prod_{i=1}^{d} b_i^{t_i}$.

The number of gates of the quantum circuit [17, Lemma 4.1] is

$$O\big(d(n \log n + d \log^3 d)\big) = O(n^{3/2} \log n)$$

(because [17, Lemma 5.6] is used $O(d)$ times, see [17, Algorithm 5.2] and the surrounding explanations). Note that this is the same as for Regev's algorithm (see [18, p5]). In our case, the number of gates is

$$O\big(d(n \log n + d^2 \log^3 d)\big) = O(n^{3/2} \log^3 n).$$

The space optimisations of Ragavan and Vaikuntanathan keep the total number of qubits for their circuit [17, Lemma 4.1] under

$$O\big(n \log n + d \log^3 d\big) = O(n \log n),$$

due to the way the qubits are used and restored in the main loop of [17, Algorithm 5.2]. In our situation, the number of qubits is

$$O\big(n \log n + d^2 \log^3 d\big) = O(n \log^3 n).$$

[10] There are extra details pertaining to the precise use of qubits (e.g. restoring the ancilla qubits to $|0\rangle$), but these will be the same in our setup.

In summary, our final quantum circuit has $O(n^{3/2}\log^3 n)$ gates and $O(n\log^3 n)$ qubits. As noted in [7], the fact that the two elements $g$ and $y$ are not small ($g$ and $y$ can be as large as $N$, as opposed to the $b_i$'s) does not affect these complexity bounds. □

**Remark 3.4.** In the proof of Theorem 1.2, we used the naive Lemma 3.2 to convert a classical circuit into a quantum one. As mentioned in Remark 1.3, it is possible to save a factor of $\log n$ in the space cost for Theorems 1.1 and 1.2 by reusing certain qubits when performing the quantum computation corresponding to Lemma 3.1.

3.3. **Factoring integers.** In this section, we prove Theorem 1.1.

**Lemma 3.5.** *Let $N > 1$ be an odd integer with at least two distinct prime factors. Let $\mathbf{x}$ be a random variable, uniformly distributed in $(\mathbb{Z}/N\mathbb{Z})^\times$. Then $\langle \mathbf{x}\rangle$ contains a non-trivial square root of 1 modulo $N$ with probability $\geqslant 1/2$.*

*Proof.* This elementary number-theoretic fact is well-known – it was already needed for Shor's algorithm [20]. See [16, Appendix A4.3] for a detailed proof. □

**Theorem 1.1.** *There is a quantum circuit having $O(n^{3/2}\log^3 n)$ quantum gates and $O(n\log^3 n)$ qubits with the following property. There is a classical randomised polynomial-time algorithm that solves the factoring problem*

$$\textbf{Input}: \text{ a composite integer } N \leqslant 2^n$$
$$\textbf{Output}: \text{ a non-trivial divisor of } N$$

*using $O(\sqrt{n})$ calls to this quantum circuit, and succeeds with probability $\Theta(1)$.*

*Proof of Theorem 1.1.* We can assume that $N$ is odd and not a perfect prime power, as otherwise it is easy to factor $N$ in polynomial time with a classical computer (see [16, Exercise 5.17]).

Let $d := \lceil \sqrt{\log N}\rceil$ and $X := d^{10^3 d}$. The probabilistic algorithm to find a non-trivial divisor of $N$ goes as follows.

(1) Generate $d$ primes $b_1, \dots, b_d$ independently and uniformly at random in the set of primes $\leqslant X$ not dividing $N$. Sample another integer $x$ uniformly chosen in $(\mathbb{Z}/N\mathbb{Z})^\times$. As in the proof of Theorem 1.2, this can be done classically in polynomial time with probability $1 - O(1/N)$.

(2) Use the algorithm of Ekerå and Gärtner [7] to obtain a basis for the lattice

$$\mathcal{L}_{\underline{b},x} := \Big\{(e_1,\dots,e_d,f)\in\mathbb{Z}^{d+1} \ : \ \Big(\prod_{i=1}^d b_i^{e_i}\Big)x^f \equiv 1 \pmod N\Big\}.$$

This involves making $O(d)$ calls to a quantum circuit, followed by polynomial-time classical post-processing. By Theorem 2.18, $\mathcal{L}_{\underline{b},x}$ has a basis of vectors of Euclidean norm $\ll e^{42d}$ with probability $1 - O(d^{-d})$, which means that this step is guaranteed to work with probability $\Theta(1)$.

(3) Using the short basis for $\mathcal{L}_{\underline{b},x}$ computed in the previous step, find the vector of the form $(0,\dots,0,r)$ in $\mathcal{L}_{\underline{b},x}$ with $r \geqslant 1$ as small as possible. This involves solving a linear system with integer coefficients, which can be done efficiently as in the proof of Theorem 1.2. This integer $r$ is the order of $x$ in $(\mathbb{Z}/N\mathbb{Z})^\times$. By Lemma 3.5, with probability $\geqslant 1/2$, this order $r$ will be even and the element $x^{r/2}$ will be a non-trivial square root of 1 modulo $N$. Hence, $N$ divides the product $(x^{r/2}-1)(x^{r/2}+1)$ but neither term individually, which implies that $\gcd(N,\, x^{r/2} - 1 \bmod N)$ is a non-trivial divisor of $N$.

The analysis of this algorithm is identical to the corresponding part of the proof of Theorem 1.2. □

## Appendix A. Bounds for character sums over primes

In this appendix, we prove Proposition 2.9. We start by stating the truncated explicit formula for $L(s,\chi)$.

**Lemma A.1.** *Let $q \geqslant 2$. Let $\chi$ be a non-principal Dirichlet character modulo $q$. For $x \geqslant T \geqslant 2$, we have*

$$\sum_{n\leqslant x}\chi(n)\Lambda(n) = -\sum_{\substack{\rho=\beta+i\gamma\\ |\gamma|\leqslant T}}\frac{x^\rho-1}{\rho} + O\left(\frac{x\log^2(xq)}{T}\right)$$

*where the sum runs over all non-trivial zeros $\rho$ of $L(s,\chi)$ with multiplicity.*

*Proof.* This is [14, Theorem 11.3]. □

We will also need the following standard bound for the number of zeros of $L(s,\chi)$ in the critical strip at some height $t$.

**Lemma A.2.** *Let $q \geqslant 1$ and let $\chi$ be a Dirichlet character modulo $q$. Let $t \in \mathbb{R}$. The number of zeros $\rho = \beta + i\gamma$ of $L(s,\chi)$ in the rectangle $0 \leqslant \beta \leqslant 1$, $t \leqslant \gamma \leqslant t+1$ is $\ll \log(q(|t|+2))$, where zeros are counted with multiplicity.*

*Proof.* This is [15, Theorem 10.17]. □

*Proof of Proposition 2.9.* By Lemmas A.1 and A.2, choosing $T = x^{1-\alpha}-1$, we have

$$\sum_{n\leqslant x}\chi(n)\Lambda(n) \ll \log(qx)\sum_{\substack{t\in\mathbb{Z}\\ |t|\leqslant x^{1-\alpha}-1}}\max_{\substack{\rho=\beta+i\gamma\\ |\gamma-t|\leqslant 1/2}}\left|\frac{x^\rho-1}{\rho}\right| + x^\alpha\log^2(qx),$$

where the maximum is over all zeros of $L(s,\chi)$ in the specified region. Since $|x^\rho| \leqslant x^\alpha$ for all zeros $\rho = \beta+i\gamma$ with imaginary part $|\gamma| \leqslant x^{1-\alpha}$ by our zero-free rectangle assumption, we can bound

$$\left|\frac{x^\rho-1}{\rho}\right| \ll \frac{x^\alpha}{1+|\gamma|}$$

(using for example that $\frac{x^\rho-1}{\rho} = \int_1^x t^{\rho-1}dt$ for $|\gamma|<1$). Thus, we obtain

$$\sum_{n\leqslant x}\chi(n)\Lambda(n) \ll \log(qx)x^\alpha\sum_{|t|\leqslant x}\frac{1}{1+|t|} + x^\alpha\log^2(qx) \ll x^\alpha\log^2(qx).$$

Discarding perfect prime powers, which contribute $O(x^{1/2}\log x)$, and using partial summation, we get

$$\sum_{p\leqslant x}\chi(p) \ll \frac{x^\alpha\log^2(qx)}{(1-\alpha)\log x}.$$

We may assume that $1-\alpha \geqslant \frac{1}{\log x}$, as otherwise Proposition 2.9 is trivial. If this is the case, we conclude that

$$\frac{1}{\pi(x)}\sum_{p\leqslant x}\chi(p) \ll \frac{x^\alpha\log^2(qx)}{\pi(x)} \ll x^{-(1-\alpha)}\log^3(qx)$$

as claimed. □

## References


[1] Manindra Agrawal, Neeraj Kayal, and Nitin Saxena, *Primes is in P*, Annals of Mathematics **160** (2004), 781–793.
[2] Nesmith C. Ankeny, *The least quadratic non residue*, Annals of mathematics (1952), 65–72.
[3] Erwin H. Bareiss, *Sylvester's identity and multistep integer-preserving Gaussian elimination*, Mathematics of Computation **22** (1968), no. 103, 565–578.
[4] Daniel J. Bernstein, Johannes Buchmann, and Erik Dahmen, *Post-quantum cryptography*, Springer, 2009.
[5] David A. Burgess, *The distribution of quadratic residues and non-residues*, Mathematika **4** (1957), no. 2, 106–112.
[6] Whitfield Diffie and Martin E. Hellman, *New directions in cryptography*, IEEE Transactions on Information Theory **22** (1976), no. 6, 644–654.
[7] Martin Ekerå and Joel Gärtner, *Extending regev's factoring algorithm to compute discrete logarithms*, preprint arXiv:2311.05545 (2023).
[8] Craig Gidney, *Asymptotically efficient quantum Karatsuba multiplication*, preprint arXiv:1904.07356 (2019).
[9] Andrew Granville and Kannan Soundararajan, *Large character sums*, Journal of the American Mathematical Society **14** (2001), no. 2, 365–397.
[10] David Harvey and Joris van der Hoeven, *Integer multiplication in time $O(n \log n)$*, Annals of Mathematics **193** (2021), no. 2, 563–617.
[11] Matti Jutila, *On Linnik's constant*, Mathematica Scandinavica **41** (1977), no. 1, 45–62.
[12] Gregory D. Kahanamoku-Meyer and Norman Y. Yao, *Fast quantum integer multiplication with zero ancillas*, preprint arXiv:2403.18006 (2024).
[13] Burton S. Kaliski Jr, *Targeted Fibonacci exponentiation*, preprint arXiv:1711.02491 (2017).
[14] Dimitris Koukoulopoulos, *The distribution of prime numbers*, vol. 203, American Mathematical Society, 2019.
[15] Hugh L. Montgomery and Robert C. Vaughan, *Multiplicative number theory I: classical theory*, Cambridge Studies in Advanced Mathematics, Cambridge University Press, 2006.
[16] Michael A. Nielsen and Isaac L. Chuang, *Quantum computation and quantum information*, Cambridge University Press, 2010.
[17] Seyoon Ragavan and Vinod Vaikuntanathan, *Optimizing space in Regev's factoring algorithm*, preprint arXiv:2310.00899 (2023).
[18] Oded Regev, *An efficient quantum factoring algorithm*, preprint arXiv:2308.06572 (2023).
[19] Ronald L. Rivest, Adi Shamir, and Leonard Adleman, *A method for obtaining digital signatures and public-key cryptosystems*, Communications of the ACM **21** (1978), no. 2, 120–126.
[20] Peter W. Shor, *Algorithms for quantum computation: discrete logarithms and factoring*, Proceedings 35th Annual Symposium on Foundations of Computer Science, IEEE, 1994, pp. 124–134.
[21] ———, *Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer*, SIAM Journal on Computing **26** (1997), no. 5, 1484–1509.
[22] Terence Tao and Van H. Vu, *Additive combinatorics*, vol. 105, Cambridge University Press, 2006.



Mathematical Institute, University of Oxford.

*Email address*: cedric.pilatte@maths.ox.ac.uk