

Quantum Counting

Gilles Brassard^{1*}, Peter Høyer^{2**}, and Alain Tapp^{1***}

¹ Université de Montréal, {brassard,tappa}@iro.umontreal.ca

² Odense University, u2pi@imada.ou.dk

Abstract. We study some extensions of Grover's quantum searching algorithm. First, we generalize the Grover iteration in the light of a concept called amplitude amplification. Then, we show that the quadratic speedup obtained by the quantum searching algorithm over classical brute force can still be obtained for a large family of search problems for which good classical heuristics exist. Finally, as our main result, we combine ideas from Grover's and Shor's quantum algorithms to perform approximate counting, which can be seen as an amplitude estimation process.

1 Introduction

Quantum computing is a field at the junction of theoretical modern physics and theoretical computer science. Practical experiments involving a few quantum bits have been successfully performed, and much progress has been achieved in quantum information theory, quantum error correction and fault tolerant quantum computation. Although we are still far from having desktop quantum computers in our offices, the quantum computational paradigm could soon be more than mere theoretical exercise [5, and references therein].

The discovery by Peter Shor [11] of a polynomial-time quantum algorithm for factoring and computing discrete logarithms was a major milestone in the history of quantum computing. Another significant result is Lov Grover's quantum search algorithm [9]. Grover's algorithm does not solve **NP**-complete problems in polynomial time, but the wide range of its applications compensates for this.

The search problem and Grover's iteration are reviewed in Section 2. It was already implicit in [6] that the heart of Grover's algorithm can be viewed as an amplitude amplification process. Here, we develop this viewpoint and obtain a more general algorithm.

When the structure in a search problem cannot be exploited, any quantum algorithm requires a computation time at least proportional to the square root of the time taken by brute-force classical searching [2]. In practice, the structure of

* Supported in part by Canada's NSERC, Québec's FCAR and the Canada Council.

** Supported in part by the ESPRIT Long Term Research Programme of the EU under project number 20244 (ALCOM-IT). Research carried out while this author was at the Université de Montréal.

*** Supported in part by postgraduate fellowships from FCAR and NSERC.

the search problem can usually be exploited, yielding deterministic or heuristic algorithms that are much more efficient than brute force would be. In Section 3, we study a vast family of heuristics for which we show how to adapt the quantum search algorithm to preserve quadratic speedup over classical techniques.

In Section 4, we present, as our main result, a quantum algorithm to perform counting. This is the problem of counting the number of elements that fulfill some specific requirements, instead of merely finding such an element. Our algorithm builds on both Grover’s iteration [9] as described in [3] and the quantum Fourier transform as used in [11]. The accuracy of the algorithm depends on the amount of time one is willing to invest. As Grover’s algorithm is a special case of the amplitude amplification process, our counting algorithm can also be viewed as a special case of the more general process of *amplitude estimation*.

We assume in this paper that the reader is familiar with basic notions of quantum computing [1, 4].

2 Quantum Amplitude Amplification

Consider the following search problem: Given a Boolean function $F : X \rightarrow \{0, 1\}$ defined on some finite domain X , find an input $x \in X$ for which $F(x) = 1$, provided such an x exists. We assume that F is given as a black box, so that it is not possible to obtain knowledge about F by any other means than evaluating it on points in its domain. The best classical strategy is to evaluate F on random elements of X . If there is a unique $x_0 \in X$ on which F takes value 1, this strategy evaluates F on roughly half the elements of the domain in order to determine x_0 . By contrast, Grover [9] discovered a quantum algorithm that only requires an expected number of evaluations of F in the order of \sqrt{N} , where $N = |X|$ denotes the cardinality of X .

It is useful for what follows to think of the above-mentioned classical strategy in terms of an algorithm that keeps boosting the probability of finding x_0 . The algorithm evaluates F on new inputs, until it eventually finds the unique input x_0 on which F takes value 1. The probability that the algorithm stops after exactly j evaluations of F is $1/N$ ($1 \leq j \leq N - 2$), and thus we can consider that each evaluation boosts the probability of success by an additive amount of $1/N$.

Intuitively, the quantum analog of boosting the probability of success would be to boost the *amplitude* of being in a certain subspace of a Hilbert space, and indeed the algorithm found by Grover can be seen as working by that latter principle [9, 3]. As discovered by Brassard and Høyer [6], the idea of amplifying the amplitude of a subspace is a technique that applies in general. Following [6], we refer to this as *amplitude amplification*, and describe the technique below. For this, we require the following notion, which we shall use throughout the rest of this section.

Let $|\mathcal{T}\rangle$ be any pure state of a joint quantum system \mathcal{H} . Write $|\mathcal{T}\rangle$ as a superposition of orthonormal states according to the state of the first subsystem:

$$|\mathcal{T}\rangle = \sum_{i \in \mathbb{Z}} x_i |i\rangle |\mathcal{T}_i\rangle$$

so that only a finite number of the states $|i\rangle|T_i\rangle$ have nonzero amplitude x_i .

Every Boolean function $\chi : \mathbb{Z} \rightarrow \{0, 1\}$ induces two orthogonal subspaces of \mathcal{H} , allowing us to rewrite $|T\rangle$ as follows:

$$|T\rangle = |T^a\rangle + |T^b\rangle = \sum_{i \in \chi^{-1}(1)} x_i |i\rangle |T_i\rangle + \sum_{i \in \chi^{-1}(0)} x_i |i\rangle |T_i\rangle. \quad (1)$$

We say that a state $|i\rangle|\cdot\rangle$ is *good* if $\chi(i) = 1$, and otherwise it is *bad*. Thus, we have that $|T^a\rangle$ denotes the projection of $|T\rangle$ onto the subspace spanned by the good states, and similarly $|T^b\rangle$ is the projection of $|T\rangle$ onto the subspace spanned by the bad states. Let $a_T = \langle T^a | T^a \rangle$ denote the probability that measuring $|T\rangle$ produces a good state, and similarly let $b_T = \langle T^b | T^b \rangle$. Since $|T^a\rangle$ and $|T^b\rangle$ are orthogonal, we have $a_T + b_T = 1$.

Let \mathcal{A} be any quantum algorithm that acts on \mathcal{H} and uses no measurements. The heart of amplitude amplification is the following operator [6]

$$\mathbf{Q} = \mathbf{Q}(\mathcal{A}, \chi, \phi, \varphi) = -\mathcal{A} \mathbf{S}_0^\phi \mathcal{A}^{-1} \mathbf{S}_\chi^\varphi. \quad (2)$$

Here, ϕ and φ are complex numbers of unit norm, and operator \mathbf{S}_χ^φ conditionally changes the phase by a factor of φ :

$$|i\rangle|\cdot\rangle \mapsto \begin{cases} \varphi |i\rangle|\cdot\rangle & \text{if } \chi(i) = 1 \\ |i\rangle|\cdot\rangle & \text{if } \chi(i) = 0. \end{cases}$$

Further, \mathbf{S}_0^ϕ changes the phase of a state by a factor of ϕ if and only if the first register holds a zero. The operator \mathbf{Q} is a generalization of the iteration applied by Grover in his original quantum searching paper [9]. It was first used in [6] to obtain an exact quantum polynomial-time algorithm for Simon's problem. It is well-defined since we assume that \mathcal{A} uses no measurements and, therefore, \mathcal{A} has an inverse.

Denote the complex conjugate of λ by λ^* . It is easy to show the following lemma by a few simple rewritings.

Lemma 1. *Let $|T\rangle$ be any superposition. Then*

$$\mathcal{A} \mathbf{S}_0^\phi \mathcal{A}^{-1} |T\rangle = |T\rangle - (1 - \phi) \langle T | \mathcal{A} | \mathbf{0} \rangle^* \mathcal{A} | \mathbf{0} \rangle.$$

By factorizing \mathbf{Q} as $(\mathcal{A} \mathbf{S}_0^\phi \mathcal{A}^{-1})(-\mathbf{S}_\chi^\varphi)$, the next lemma follows.

Lemma 2. *Let $|T\rangle = |T^a\rangle + |T^b\rangle$ be any superposition. Then*

$$\mathbf{Q} |T^a\rangle = -\varphi |T^a\rangle + \varphi(1 - \phi) \langle T^a | \mathcal{A} | \mathbf{0} \rangle^* \mathcal{A} | \mathbf{0} \rangle \quad (3)$$

$$\mathbf{Q} |T^b\rangle = -|T^b\rangle + (1 - \phi) \langle T^b | \mathcal{A} | \mathbf{0} \rangle^* \mathcal{A} | \mathbf{0} \rangle. \quad (4)$$

In particular, letting $|T\rangle$ be $\mathcal{A} | \mathbf{0} \rangle = |\Psi^a\rangle + |\Psi^b\rangle$ implies that the subspace spanned by $|\Psi^a\rangle$ and $|\Psi^b\rangle$ is invariant under the action of \mathbf{Q} .

Lemma 3. Let $\mathcal{A}|0\rangle = |\Psi\rangle = |\Psi^a\rangle + |\Psi^b\rangle$. Then

$$\mathbf{Q}|\Psi^a\rangle = \varphi((1-\phi)a-1)|\Psi^a\rangle + \varphi(1-\phi)a|\Psi^b\rangle \quad (5)$$

$$\mathbf{Q}|\Psi^b\rangle = -((1-\phi)a+\phi)|\Psi^b\rangle + (1-\phi)(1-a)|\Psi^a\rangle, \quad (6)$$

where $a = \langle\Psi^a|\Psi^a\rangle$.

From Lemmas 2 and 3 it follows that, for any vector $|\mathcal{T}\rangle = |\mathcal{T}^a\rangle + |\mathcal{T}^b\rangle$, the subspace spanned by the set $\{|\mathcal{T}^a\rangle, |\mathcal{T}^b\rangle, |\Psi^a\rangle, |\Psi^b\rangle\}$ is invariant under the action of \mathbf{Q} . By setting $\phi = \varphi = -1$, we find the following much simpler expressions.

Lemma 4. Let $\mathcal{A}|0\rangle = |\Psi\rangle = |\Psi^a\rangle + |\Psi^b\rangle$, and let $\mathbf{Q} = \mathbf{Q}(\mathcal{A}, \chi, -1, -1)$. Then

$$\mathbf{Q}|\Psi^a\rangle = (1-2a)|\Psi^a\rangle - 2a|\Psi^b\rangle \quad (7)$$

$$\mathbf{Q}|\Psi^b\rangle = (1-2a)|\Psi^b\rangle + 2b|\Psi^a\rangle, \quad (8)$$

where $a = \langle\Psi^a|\Psi^a\rangle$ and $b = 1-a = \langle\Psi^b|\Psi^b\rangle$.

The recursive formulae defined by Equations 7 and 8 were solved in [3], and their solution is given in the following theorem. The general cases defined by Equations 3–6 have similar solutions, but we shall not need them in what follows.

Theorem 1 (Amplitude Amplification—simple case). Let $\mathcal{A}|0\rangle = |\Psi\rangle = |\Psi^a\rangle + |\Psi^b\rangle$, and let $\mathbf{Q} = \mathbf{Q}(\mathcal{A}, \chi, -1, -1)$. Then, for all $j \geq 0$,

$$\mathbf{Q}^j \mathcal{A}|0\rangle = k_j |\Psi^a\rangle + \ell_j |\Psi^b\rangle,$$

where

$$k_j = \frac{1}{\sqrt{a}} \sin((2j+1)\theta) \quad \text{and} \quad \ell_j = \frac{1}{\sqrt{1-a}} \cos((2j+1)\theta),$$

and where θ is defined so that $\sin^2(\theta) = a = \langle\Psi^a|\Psi^a\rangle$ and $0 \leq \theta \leq \pi/2$.

Theorem 1 yields a method for boosting the success probability a of a quantum algorithm \mathcal{A} . Consider what happens if we apply \mathcal{A} on the initial state $|0\rangle$ and then measure the system. The probability that the outcome is a good state is a . If, instead of applying \mathcal{A} , we apply operator $\mathbf{Q}^m \mathcal{A}$ for some integer $m \geq 1$, then our success probability is given by $ak_m^2 = \sin^2((2m+1)\theta)$. Therefore, to obtain a high probability of success, we want to choose integer m such that $\sin^2((2m+1)\theta)$ is close to 1. Unfortunately, our ability to choose m wisely depends on our knowledge about θ , which itself depends on a . The two extreme cases are when we know the exact value of a , and when we have no prior knowledge about a whatsoever.

Suppose the value of a is known. If $a > 0$, then by letting $m = \lfloor \pi/4\theta \rfloor$, we have that $ak_m^2 \geq 1-a$, as shown in [3]. The next theorem is immediate.

Theorem 2 (Quadratic speedup). *Let \mathcal{A} be any quantum algorithm that uses no measurements, and let $\chi : \mathbb{Z} \rightarrow \{0, 1\}$ be any Boolean function. Let the initial success probability a and angle θ be defined as in Theorem 1. Suppose $a > 0$ and set $m = \lfloor \pi/4\theta \rfloor$. Then, if we compute $\mathbf{Q}^m \mathcal{A}|0\rangle$ and measure the system, the outcome is good with probability at least $\max(1 - a, a)$.*

This theorem is often referred to as a quadratic speedup, or the square-root running-time result. The reason for this is that if an algorithm \mathcal{A} has success probability $a > 0$, then after an expected number of $1/a$ applications of \mathcal{A} , we will find a good solution. Applying the above theorem reduces this to an expected number of at most $(2m + 1)/(1 - a) \in \Theta(\sqrt{1/a})$ applications of \mathcal{A} and its inverse.

Suppose the value of a is known and that $0 < a < 1$. Theorem 2 allows us to find a good solution with probability at least $\max(1 - a, a)$. A natural question to ask is whether it is possible to improve this to certainty, still given the value of a . It turns out that the answer is positive. This is unlike classical computers, where no such general de-randomization technique is known. We now describe two optimal methods for obtaining this, but other approaches are possible.

The first method is by applying amplitude amplification, not on the original algorithm \mathcal{A} , but on a slightly modified version of it. If $\tilde{m} = \pi/4\theta - 1/2$ is an integer, then we would have $\ell_{\tilde{m}} = 0$, and we would succeed with certainty. In general, $m_0 = \lceil \tilde{m} \rceil$ iterations is a fraction of 1 iteration too many, but we can compensate for that by choosing $\theta_0 = \pi/(4m_0 + 2)$, an angle slightly smaller than θ . Any quantum algorithm that succeeds with probability a_0 such that $\sin^2(\theta_0) = a_0$, will succeed with certainty after m_0 iterations of amplitude amplification. Given \mathcal{A} and its initial success probability a , it is easy to construct a new quantum algorithm that succeeds with probability $a_0 \leq a$: Let \mathcal{B} denote the quantum algorithm that takes a single qubit in the initial state $|0\rangle$ and rotates it to the superposition $\sqrt{1 - a_0/a}|0\rangle + \sqrt{a_0/a}|1\rangle$. Apply both \mathcal{A} and \mathcal{B} , and define a good solution as one in which \mathcal{A} produces a good solution, and the outcome of \mathcal{B} is the state $|1\rangle$.

The second method is to slow down the speed of the very last iteration. First, apply $m_0 = \lceil \tilde{m} \rceil$ iterations of amplitude amplification with $\phi = \varphi = -1$. Then, if $m_0 < \tilde{m}$, apply one more iteration with complex phase-shifts ϕ and φ satisfying $\ell_{m_0}^2 = 2a(1 - \text{Re}(\phi))$ and so that $\varphi(1 - \phi)ak_{m_0} - ((1 - \phi)a + \phi)\ell_{m_0}$ vanishes. Going through the algebra and applying Lemma 3 shows that this produces a good solution with certainty. For the case $m_0 = 0$, this second method was independently discovered by Chi and Kim [7].

Suppose now that the value of a is not known. In Section 4, we discuss techniques for finding a good estimate of a , after which one then can apply a weakened version of Theorem 2 to find a good solution. Another idea is to try to find a good solution without prior computation of an estimate of a . Within that approach, by adapting the ideas in Section 4 in [3] (Section 6 in its final version), we can still obtain a quadratic speedup.

Theorem 3 (Quadratic speedup without knowing a). *Let \mathcal{A} be any quantum algorithm that uses no measurements, and let $\chi : \mathbb{Z} \rightarrow \{0, 1\}$ be any Boolean*

function. Let the initial success probability a of \mathcal{A} be defined as in Theorem 1. Then there exists a quantum algorithm that finds a good solution using an expected number of $\Theta(\sqrt{1/a})$ applications of \mathcal{A} and its inverse if $a > 0$, and otherwise runs forever.

By applying this theorem to the searching problem defined in the first paragraph of this section, we obtain the following result from [3], which itself is a generalization of the work by Grover [9].

Corollary 1. *Let $F : X \rightarrow \{0, 1\}$ be any Boolean function defined on a finite set X . Then there exists a quantum algorithm **Search** that finds an $x \in X$ such that $F(x) = 1$ using an expected number of $\Theta(\sqrt{|X|/t})$ evaluations of F , provided such an x exists, and otherwise runs forever. Here $t = |\{x \in X \mid F(x) = 1\}|$ denotes the cardinality of the preimage of 1.*

Proof. Apply Theorem 3 with $\chi = F$ and \mathcal{A} being any unitary transformation that maps $|0\rangle$ to $\frac{1}{\sqrt{|X|}} \sum_{x \in X} |x\rangle$, such as the Walsh–Hadamard transform. \square

3 Quantum Heuristics

If function F has no useful structure, then quantum algorithm **Search** will be more efficient than any classical (deterministic or probabilistic) algorithm. In sharp contrast, if some useful information is known about the function, then some classical algorithm might be very efficient. Useful information might be clear mathematical statements or intuitive information stated as a probability distribution of the likelihood of x being a solution. The information we have about F might also be expressed as an efficient classical heuristic to find a solution. In this section, we address the problem of heuristics.

Search problems, and in particular **NP** problems, are often very difficult to solve. For many **NP**-complete problems, practical algorithms are known that are more efficient than brute force search on the average: they take advantage of the problem’s structure and especially of the input distribution. Although in general very few theoretical results exist about the efficiency of heuristics, they are very efficient in practice.

We concentrate on a large but simple family of heuristics that can be applied to search problems. Here, by heuristics, we mean a probabilistic algorithm running in polynomial time that outputs what one is searching for with some nonzero probability. Our goal is to apply Grover’s technique for heuristics in order to speed them up, in the same way that Grover speeds up black-box search, without making things too complicated.

More formally, suppose we have a family \mathcal{F} of functions such that each $F \in \mathcal{F}$ is of the form $F : X \rightarrow \{0, 1\}$. A *heuristic* is a function $G : \mathcal{F} \times R \rightarrow X$, for an appropriate finite set R . For every function $F \in \mathcal{F}$, let $t_F = |F^{-1}(1)|$ and $h_F = |\{r \in R \mid F(G(F, r)) = 1\}|$. We say that the heuristic is *efficient* for a given F if $h_F/|R| > t_F/|X|$ and the heuristic is *good* in general if

$$\mathbb{E}_{\mathcal{F}} \left(\frac{h_F}{|R|} \right) > \mathbb{E}_{\mathcal{F}} \left(\frac{t_F}{|X|} \right) .$$

Here $E_{\mathcal{F}}$ denotes the expectation over all F according to some fixed distribution. Note that for some F , h_F might be small but repeated uses of the heuristic, with seeds r uniformly chosen in R , will increase the probability of finding a solution.

Theorem 4. *Given a search problem F chosen in a family \mathcal{F} according to some distribution, if using a heuristic G , a solution to F is found in expected time T then, using a quantum computer, a solution can be found in expected time in $O(\sqrt{T})$.*

Proof. We simply combine the quantum algorithm **Search** with the heuristic G . Let $G'(r) = F(G(F, r))$, clearly $x = G(F, \mathbf{Search}(G'))$ is such that $F(x) = 1$. Thus, by Corollary 1, for each function $F \in \mathcal{F}$, we have an expected running time of $\Theta(\sqrt{|R|/h_F})$. Let P_F denote the probability that F occurs. Then $\sum_{F \in \mathcal{F}} P_F = 1$, and we have that the expected running time is of order $\sum_{F \in \mathcal{F}} \sqrt{|R|/h_F} P_F$, which can be rewritten as

$$\sum_{F \in \mathcal{F}} \sqrt{\frac{|R|}{h_F}} P_F \sqrt{P_F} \leq \left(\sum_{F \in \mathcal{F}} \frac{|R|}{h_F} P_F \right)^{1/2} \left(\sum_{F \in \mathcal{F}} P_F \right)^{1/2} = \left(\sum_{F \in \mathcal{F}} \frac{|R|}{h_F} P_F \right)^{1/2},$$

by Cauchy–Schwarz’s inequality. \square

4 Approximate Counting

In this section, we do not concentrate on finding one solution, but rather on counting them. For this, we complement Grover’s iteration [9] using techniques inspired by Shor’s quantum factoring algorithm [11].

Counting Problem: Given a Boolean function F defined on some finite set $X = \{0, \dots, N-1\}$, find or approximate $t = |F^{-1}(1)|$.

Before we proceed, here is the basic intuition. From Section 2 it follows that, in Grover’s algorithm, the amplitude of the set $F^{-1}(1)$, as well as the amplitude of the set $F^{-1}(0)$, varies with the number of iterations according to a periodic function. We also note that the period (frequency) of this association is in *direct* relation with the sizes of these sets. Thus, estimating their common period using Fourier analysis will give us useful information on the sizes of those two sets. Since the period will be the same if $F^{-1}(1)$ has cardinality t , as if $F^{-1}(1)$ has cardinality $N - t$, we will in the rest of this section assume that $t \leq N/2$.

The quantum algorithm **Count** we give to solve this problem has two parameters: the function F given as a black box and an integer P that will determine the precision of our estimate, as well as the time taken by the algorithm. For simplicity, we assume that P and N are powers of 2, but this is not essential. Our algorithm is based on the following two unitary transformations:

$$\begin{aligned} \mathbf{C}_F : |m\rangle \otimes |\Psi\rangle &\rightarrow |m\rangle \otimes (\mathbf{G}_F)^m |\Psi\rangle \\ \mathbf{F}_P : |k\rangle &\rightarrow \frac{1}{\sqrt{P}} \sum_{l=0}^{P-1} e^{2\pi i k l / P} |l\rangle. \end{aligned}$$

Here $\imath = \sqrt{-1}$ and $\mathbf{G}_F = \mathbf{Q}(\mathbf{W}, F, -1, -1)$ denotes the iteration originally used by Grover [9], where \mathbf{W} denotes the Walsh–Hadamard transform on n qubits that maps $|0\rangle$ to $2^{-n/2} \sum_{i=0}^{2^n-1} |i\rangle$.

In order to apply \mathbf{C}_F even if its first argument is in a quantum superposition, it is necessary to have an upper bound on the value of m , which is the purpose of parameter P . Thus, unitary transformation \mathbf{C}_F performs exactly P Grover’s iterations so that P evaluations of F are required. The quantum Fourier transform can be efficiently implemented (see [11] for example).

Count(F, P)

1. $|\Psi_0\rangle \leftarrow \mathbf{W} \otimes \mathbf{W} |0\rangle|0\rangle$
2. $|\Psi_1\rangle \leftarrow \mathbf{C}_F |\Psi_0\rangle$
3. $|\Psi_2\rangle \leftarrow |\Psi_1\rangle$ after the second register is measured (*optional*)
4. $|\Psi_3\rangle \leftarrow \mathbf{F}_P \otimes \mathbf{I} |\Psi_2\rangle$
5. $\tilde{f} \leftarrow \text{measure } |\Psi_3\rangle$ (if $\tilde{f} > P/2$ then $\tilde{f} \leftarrow (P - \tilde{f})$)
6. output: $N \sin^2(\tilde{f}\pi/P)$ (and \tilde{f} if needed)

The following theorem tells us how to make proper use of algorithm **Count**.

Theorem 5. *Let $F : \{0, \dots, N-1\} \rightarrow \{0, 1\}$ be a Boolean function, $t = |F^{-1}(1)| \leq N/2$ and \tilde{t} be the output of **Count**(F, P) with $P \geq 4$, then*

$$|t - \tilde{t}| < \frac{2\pi}{P} \sqrt{tN} + \frac{\pi^2}{P^2} N$$

with probability at least $8/\pi^2$.

Proof. Let us follow the state through the algorithm using notation from Section 2.

$$|\Psi_0\rangle = \frac{1}{\sqrt{PN}} \sum_{m=0}^{P-1} \sum_{x=0}^{N-1} |m\rangle|x\rangle$$

$$|\Psi_1\rangle = \frac{1}{\sqrt{P}} \sum_{m=0}^{P-1} |m\rangle \left(k_m \sum_{x \in F^{-1}(1)} |x\rangle + \ell_m \sum_{x \in F^{-1}(0)} |x\rangle \right).$$

We introduced Step 3 to make it intuitively clear to the reader why the Fourier transform in Step 4 gives us what we want. The result of this measurement is not used in the algorithm and this is why it is optional: the final outcome would be the same if Step 3 were not performed. Without loss of generality, assume that the state x observed in the second register is such that $F(x) = 1$. Then by replacing k_m by its definition we obtain

$$|\Psi_2\rangle = \alpha \sum_{m=0}^{P-1} \sin((2m+1)\theta) |m\rangle, \tag{9}$$

where α is a normalization factor that depends on θ .

Let

$$f = P\theta/\pi. \quad (10)$$

In Step 4, we apply the Fourier transform on a sine (cosine) of period f and phase shift θ . From $\sin^2(\theta) = t/N$ we conclude that $\theta \leq \pi/2$ and $f \leq P/2$. After we apply the Fourier transform, the state $|\Psi_3\rangle$ strongly depends on f (which depends on t). If f were an integer, there would be two possibilities: either $f = 0$ (which happens if $t = 0$ or $t = N$), in which case $|\Psi_3\rangle = |0\rangle$, or $t > 0$, in which case $|\Psi_3\rangle = a|f\rangle + b|P - f\rangle$, where a and b are complex numbers of norm $1/\sqrt{2}$.

In general f is not an integer and we will obtain something more complicated. We define $f^- = \lfloor f \rfloor$ and $f^+ = \lfloor f + 1 \rfloor$. We still have three cases. If $1 < f < P/2 - 1$, we obtain

$$|\Psi_3\rangle = a|f^-\rangle + b|f^+\rangle + c|P - f^-\rangle + d|P - f^+\rangle + |R\rangle$$

where $|R\rangle$ is an un-normalized error term that may include some or all values other than the desirable f^- , f^+ , $P - f^-$ and $P - f^+$. The two other possibilities are $0 < f < 1$, in which case we obtain

$$|\Psi_3\rangle = a|0\rangle + b|1\rangle + c|P - 1\rangle + |R\rangle$$

or $P/2 - 1 < f < P/2$, in which case we obtain

$$|\Psi_3\rangle = a|P/2 - 1\rangle + b|P/2\rangle + c|P/2 + 1\rangle + |R\rangle.$$

In all three cases, extensive algebraic manipulation shows that the square of the norm of the error term $|R\rangle$ can be upper bounded by $2/5$,

$$\langle R|R \rangle < \frac{2}{5}.$$

In order to bound the success probability by $8/\pi^2$ (which is roughly 0.81 and therefore larger than $1 - 2/5 = 0.6$) as claimed in the statement of the Theorem, we could perform a complicated case analysis depending on whether the value x observed in Step 3 is such that $F(x) = 0$ or $F(x) = 1$. Fortunately, in the light of some recent analysis of Michele Mosca [10], which itself is based on results presented in [8], this analysis can be simplified. Since the information obtained by measuring the second register is not used, measuring it in a different basis would not change the behaviour of the algorithm. Measuring in the eigenvector basis of \mathbf{G}_F , one obtains this bound in an elegant way. Details will be provided in the final version of this paper.

Assuming that \tilde{f} has been observed at Step 5 and applying Equation 10 and the fact that $\sin(\theta) = \sqrt{t/N}$, we obtain an estimate \tilde{t} of t such that

$$|t - \tilde{t}| < \frac{2\pi}{P}\sqrt{tN} + \frac{\pi^2}{P^2}N.$$

□

Using a similar technique, it can be shown that the same quantum algorithm can also be used to perform amplitude estimation: Grover's algorithm [9] is to amplitude amplification what approximate counting is to amplitude estimation.

Theorem 6. *Replacing \mathbf{G}_F in \mathbf{C}_F of algorithm **Count** by $\mathbf{Q} = \mathbf{Q}(\mathcal{A}, \chi, -1, -1)$ and also modifying Step 6 so that the algorithm outputs $\tilde{a} = \sin^2(\tilde{f}\pi/P)$, **Count**(F, P) with $P \geq 4$ will output \tilde{a} such that*

$$|a - \tilde{a}| < \frac{2\pi}{P}\sqrt{a} + \frac{\pi^2}{P^2}$$

with probability at least $8/\pi^2$.

In Theorems 5 and 6, parameter P allows us to balance the desired accuracy of the estimate with the running time required to achieve it. We will now look at different choices for P and analyze the accuracy of the answer. To obtain t up to a few standard deviations, apply the following corollary of Theorem 5.

Corollary 2. *Given a Boolean function $F : \{0, \dots, N-1\} \rightarrow \{0, 1\}$ with t as defined above, **Count**($F, c\sqrt{N}$) outputs an estimate \tilde{t} such that*

$$|t - \tilde{t}| < \frac{2\pi}{c}\sqrt{t} + \frac{\pi^2}{c^2}$$

with probability at least $8/\pi^2$ and requires exactly $c\sqrt{N}$ evaluations of F .

The above corollary states that some accuracy can be achieved with probability $8/\pi^2$. This means that, as usual, the success probability can be boosted exponentially close to 1 by repetition. We will denote by **Maj**(k, \mathbf{Count}) an algorithm that performs k evaluations of **Count** and outputs the majority answer. To obtain an error probability smaller than $1/2^n$, one should choose k in $\Omega(n)$.

If one is satisfied in counting up to a constant relative error, it would be natural to call **Count** with $P = c\sqrt{N/t}$, but we need to use the following strategy because t is precisely what we are looking for.

CountRel(F, c)

1. $P \leftarrow 2$
2. Repeat
 - (a) $P \leftarrow 2P$
 - (b) $\tilde{f} \leftarrow \mathbf{Maj}(\Omega(\log \log N), \mathbf{Count}(F, P))$
3. Until $\tilde{f} > 1$
4. Output **Count**(F, cP)

Note that in the main loop the algorithm calls **Count** to obtain \tilde{f} and not \tilde{t} .

Corollary 3. *Given F with N and t as defined above, **CountRel**(F, c) outputs an estimate \tilde{t} such that*

$$|t - \tilde{t}| < t/c$$

with probability at least $\frac{3}{4}$, using an expected number of $\Theta((c + \log \log N)\sqrt{N/t})$ evaluations of F .

Proof. Suppose for the moment that in Step 2(b) we always obtain \tilde{f} such that $|f - \tilde{f}| < 1$. Combining this with Equation 10 we see that to obtain $\tilde{f} > 1$, we must have $P\theta/\pi > 1$. Since $\sin(\theta)^2 = t/N$, then $P > 2\sqrt{N/t}$, so, by Theorem 5, $|t - \tilde{t}| < t\frac{\pi}{c}(1 + \frac{\pi}{c})$. Thus, the core of the main loop will be performed at most $\log(2\sqrt{N/t})$ times before P is large enough. By using $\Omega(\log \log N)$ repetitive calls to **Count** in Step 2(b), we know that this will happen with sufficiently high probability, ensuring an overall success probability of at least $3/4$.

The expected number of evaluations of F follows from the fact that $\sum_{i=1}^{\log(2\sqrt{N/t})} (\log \log N) 2^i \in \Theta((\log \log N)\sqrt{N/t})$. \square

Of course, to obtain a smaller relative error, the first estimate can be used in order to call **Count** with P as large as one wishes. From Theorem 5, it is clear that by letting P be large enough, one can make the absolute error smaller than 1.

Corollary 4. *Given F with N and t as defined above, there is an algorithm requiring an expected number of $\Theta(\sqrt{tN})$ evaluations of F that outputs an estimate \tilde{t} such that $\tilde{t} = t$ with probability at least $\frac{3}{4}$ using only space linear in $\log N$.*

Proof. By Theorem 5, if $P > \pi(2 + \sqrt{6})\sqrt{tN}$, the error in the output of **Count** is likely to be smaller than $1/2$. Again we do not know t , but we already know how to estimate it. By calling first **Count**(F, \sqrt{N}) a few times, we obtain an approximation \tilde{t} such that $|t - \tilde{t}| < 2\pi\sqrt{t} + \pi^2$ with good probability. Now, assuming the first estimate was good, calling **Count**($F, 20\sqrt{\tilde{t}N}$) we obtain $\tilde{t}' = t$ with a probability of at least $8/\pi^2$. Thus, obtaining an overall success probability of at least $3/4$. \square

Note that successive applications of Grover's algorithm in which we strike out the solutions as they are found will also provide an exact count with high probability in a time in $O(\sqrt{tN})$, but at a high cost in terms of additional quantum memory, that is $\Theta(t)$.

Acknowledgements

We are grateful to Joan Boyar, Harry Buhrman, Christoph Dürr, Michele Mosca, Barbara Terhal and Ronald de Wolf for helpful comments. The third author would like to thank Mélanie Doré Boulet for her encouragements throughout the realization of this work.

References

1. BARENCO, Adriano, "Quantum physics and computers", *Contemporary Physics*, Vol. 38, 1996, pp. 357–389.
2. BENNETT, Charles H., Ethan BERNSTEIN, Gilles BRASSARD and Umesh VAZIRANI, "Strengths and weaknesses of quantum computing", *SIAM Journal on Computing*, Vol. 26, no. 5, October 1997, pp. 1510–1523.

3. BOYER, Michel, Gilles BRASSARD, Peter HØYER and Alain TAPP, "Tight bounds on quantum searching", *Proceedings of Fourth Workshop on Physics and Computation — PhysComp '96*, November 1996, pp. 36–43. Final version to appear in *Fortschritte Der Physik*.
4. BRASSARD, Gilles, "A quantum jump in computer science", in *Computer Science Today*, Jan van Leeuwen (editor), Lecture Notes in Computer Science, Vol. 1000, Springer-Verlag, 1995, pp. 1–14.
5. BRASSARD, Gilles, "New horizons in quantum information processing", *Proceedings of this ICALP Conference*, 1998.
6. BRASSARD, Gilles and Peter HØYER, "An exact quantum polynomial-time algorithm for Simon's problem", *Proceedings of Fifth Israeli Symposium on Theory of Computing and Systems — ISTCS '97*, June 1997, IEEE Computer Society Press, pp. 12–23.
7. CHI, Dong-Pyo and Jinsoo KIM, "Quantum database searching by a single query", Lecture at First NASA International Conference on Quantum Computing and Quantum Communications, Palm Springs, February 1998.
8. CLEVE, Richard, Artur EKERT, Chiara MACCHIAVELLO and Michele MOSCA, "Quantum algorithms revisited", *Proceedings of the Royal Society, London*, Vol. A354, 1998, pp. 339–354.
9. GROVER, Lov K., "Quantum mechanics helps in searching for a needle in a haystack", *Physical Review Letters*, Vol. 79, no. 2, 14 July 1997, pp. 325–328.
10. MOSCA, Michele, "Quantum computer algorithms and interferometry", Lecture at *BRICS Workshop on Algorithms in Quantum Information Processing*, Aarhus, January 1998.
11. SHOR, Peter W., "Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer", *SIAM Journal on Computing*, Vol. 26, no. 5, October 1997, pp. 1484–1509.