Trade-offs in the quantum search algorithm

Lov K. Grover*

1D435 Bell Laboratories, Lucent Technologies, 600-700 Mountain Avenue, Murray Hill, New Jersey 07974

(Received 19 December 2001; published 21 November 2002)

Quantum search has been proved to be the best possible algorithm for the exhaustive search problem in the sense that the number of queries it requires cannot be reduced. However, the number of nonquery operations, and thus the total number of operations, can be reduced. The number of nonquery unitary operations can be reduced by a factor of $\log N/\alpha \log(\log N)$ while increasing the number of queries by a factor of only $[1+(\log N)^{-\alpha}]$. For example, by choosing α to be $O(\log N/\log(\log N))$, the number of nonquery unitary operations can be reduced by 40% while increasing the number of queries by just two.

DOI: 10.1103/PhysRevA.66.052314 PACS number(s): 03.67.Lx

I. INTRODUCTION

Quantum search is a quantum-mechanical technique for searching N possibilities in only $O(\sqrt{N})$ steps. It has been proved through subtle properties of unitary transformations that the number of queries required by the algorithm is optimal [3,2]. This is usually expressed by saying that "the quantum search algorithm is the best possible algorithm for exhaustive search." It is true that the number of queries required cannot be reduced, however there is room for improvement in the total number of operations required by the algorithm. This is achieved by breaking up the nonquery transformations into bitwise operations in a way somewhat reminiscent of the techniques used to improve the sorting algorithm beyond the information theoretic limit [4].

It is shown that by slightly increasing the number of queries, the total number of operations can be reduced by a logarithmic factor. This is accomplished by making use of the amplitude amplification principle.

II. AMPLITUDE AMPLIFICATION

A few years after the invention of the quantum search algorithm, it was generalized to a much larger class of applications known as the amplitude amplification algorithms [5] (similar results are independently proved in [6]). In these algorithms, the amplitude produced in a particular state, t (t for target), by a unitary operation U when applied to an s state (s for source) can be *amplified* by successively repeating the sequence of operations: $Q=I_sU^{\dagger}I_tU$. Here I_s and I_t denote quantum transformations that selectively invert the amplitudes in the states s and t, respectively. If we start from the s state and repeat the operation sequence $I_sU^{\dagger}I_tU$, η times, followed by a single repetition of U, then the amplitude in the t state becomes approximately $2\eta U_{ts}$ (provided $\eta|U_{ts}|\ll 1$). Also, if we start from s and carry out $\pi/4|U_{ts}|$ repetitions of Q followed by a single repetition of U, we reach t with certainty. The quantum search algorithm is a particular case of this with U being the Walsh-Hadamard transformation (W) and s being the $\bar{0}$ state (the Walsh-Hadamard transformation is the operation $H=1/\sqrt{2}\begin{bmatrix}1 & 1\\ 1 & -1\end{bmatrix}$, independently applied to each qubit in the set of n qubits that were being used to represent 2^n states).

The power of the amplitude amplification technique lies in the fact that U can be *any* unitary operation. Once we can design a unitary operation (or a sequence of unitary operations) U that produces a certain amplitude in the target state, the amplitude amplification principle gives a prescription for *amplifying* this amplitude. The amount of amplification increases linearly with the number of repetitions of Q and hence the probability of detecting t goes up quadratically. For many applications, this results in a square-root speed up over the equivalent classical algorithm. In this paper, we use the amplitude amplification principle for enhancing the quantum search algorithm. This is achieved by designing a sequence of bitwise operations that produces almost the same amplitude in the t state while requiring fewer operations.

III. THE QUANTUM SEARCH ALGORITHM

As mentioned before, the quantum search algorithm is a particular case of amplitude amplification with the Walsh-Hadamard transformation being the U operation and s being the $\bar{0}$ state. For any t, $|U_{ts}|=1/\sqrt{N}$. It follows from the amplitude amplification principle that if we start from $\bar{0}$ and carry out $\pi\sqrt{N}/4$ repetitions of $-I_{\bar{0}}WI_tW$, followed by W, we reach the t state with certainty. Equivalently:

$$W\underbrace{(-I_{\bar{0}}WI_tW)\cdots(-I_{\bar{0}}WI_tW)(-I_{\bar{0}}WI_tW)(-I_{\bar{0}}WI_tW)}_{\frac{\pi\sqrt{N}}{4}\text{ repetitions}}|\bar{0}\rangle=|t\rangle.$$

Let N be the number of items being searched. Then $I_{\bar{0}}$ requires us to calculate the AND of $\log_2 N$ boolean variables which can be carried out by $\log_2 N$ C^2NOT operations. W requires $\log_2 N$ one-qubit operations since it requires only one operation per qubit. Thus the total number of additional (nonquery) qubit operations required by the algorithm is $\pi\sqrt{N}/4\times 3\times\log_2 N$ while the number of queries required is $\pi\sqrt{N}/4$. In the following section, we show how to reduce the number of additional (nonquery) qubit operations while keeping the number of queries approximately the same.

*Email address: lkgrover@bell-labs.com

1050-2947/2002/66(5)/052314(5)/$20.00

©2002 The American Physical Society

IV. INVERSION ABOUT AVERAGE

The quantum search algorithm was first presented in terms of the *inversion about average* transformation [1,7]. This paper combines the inversion about average transformation with the amplitude amplification technique to obtain a faster algorithm for exhaustive search. Before presenting the new algorithm, we first recall the inversion about average transformation.

Consider the operation sequence $(-WI_{\bar{0}}W)$. This may be written as $-W(I-2|\bar{0}\rangle\langle\bar{0}|)W$ or equivalently $(2W|\bar{0}\rangle\langle\bar{0}|W-I)$. The transformation $W|\bar{0}\rangle\langle\bar{0}|W$ can be represented as an $N\times N$ matrix with each entry equal to $1/N$, therefore each element of the transformed vector is equal to the average of all elements of the initial vector, i.e., if the ith component of the input vector $\bar{\alpha}$ is α_i, then each component of the vector $W|\bar{0}\rangle\langle\bar{0}|W\bar{\alpha}$ is α_{av}, where $\alpha_{\mathrm{av}}\equiv(1/N)\Sigma_i\alpha_i$. Hence the ith component of the transformed vector $(2W|\bar{0}\rangle\langle\bar{0}|W-I)\bar{\alpha}$ is equal to $2\alpha_{\mathrm{av}}-\alpha_i$. This may be written as $\alpha_{\mathrm{av}}-(\alpha_i-\alpha_{\mathrm{av}})$, i.e., the ith component in the transformed vector is as much below the average as the ith component in the initial vector was above the average; i.e., this transformation is an *inversion about average.*

As mentioned before, the quantum search algorithm consists of the operation sequence

$$W\underbrace{(-I_{\bar{0}}WI_tW)\cdots(-I_{\bar{0}}WI_tW)(-I_{\bar{0}}WI_tW)(-I_{\bar{0}}WI_tW)}_{\frac{\pi\sqrt{N}}{4}\text{ repetitions}}|\bar{0}\rangle.$$

This may be written as

$$\underbrace{(-WI_{\bar{0}}W)I_t\cdots(-WI_{\bar{0}}W)I_t(-WI_{\bar{0}}W)I_t}_{\frac{\pi\sqrt{N}}{4}\text{ repetitions}}W|\bar{0}\rangle.$$

This has the following interpretation.

(i) $W|\bar{0}\rangle$ creates a superposition with equal amplitude in each of N states.

(ii) I_t selectively inverts the amplitude in the target state.

Next the sequence of operations (3–4) is repeated $\pi\sqrt{N}/4$ times.

(iii) $(-WI_{\bar{0}}W)$. As described above, this is the inversion about average transformation. The average amplitude (α_{av}) is approximately equal to the amplitude of the $(N-1)$ nontarget states. Therefore, as a result of this transformation, the amplitude in the nontarget states is unaltered. Since the t state is inverted, its amplitude is below the average. As described in [1], its amplitude changes sign and its magnitude increases by $2\alpha_{\mathrm{av}}$.

(iv) I_t selectively inverts the amplitude in the target state thus undoing the sign change in (3). This prepares the system for the next inversion about average operation through which the magnitude of the amplitude in the t state is increased.

V. PARTIAL INVERSION ABOUT AVERAGE

Assume there to be n qubits. Then as described in the previous section, $-WI_{\bar{0}}W$ does an inversion about average transformation on the entire set of $N\equiv2^n$ states. Consider a set that contains m of the n qubits; denote this set by S. Define the Walsh-Hadamard transformation on S as the operation $H=1/\sqrt{2}\left[\begin{smallmatrix}1 & 1\\ 1 & -1\end{smallmatrix}\right]$, applied to each qubit in the set S, and denote this by $W^{(S)}$. Similarly define the operation $I_{\bar{0}}^{(S)}$ as the selective inversion of the state in which each qubit in S is 0.

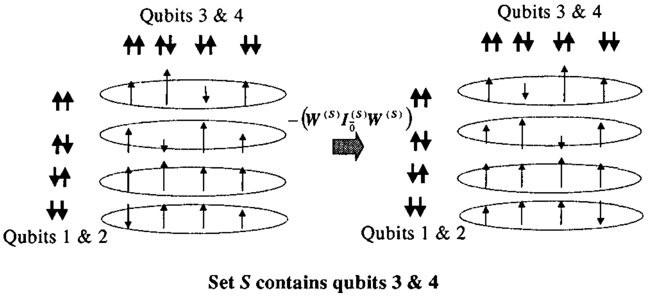
Qubits 3 & 4
Qubits 3 & 4
Qubits 1 & 2
Qubits 1 & 2
$-(W^{(S)}I_{\bar{0}}^{(S)}W^{(S)})$
Set S contains qubits 3 & 4

FIG. 1. $-W^{(S)}I_{\bar{0}}^{(S)}W^{(S)}$ performs an inversion about average in each of the four subsets of states. The four subsets are defined by the condition that the qubits *not* in S stay fixed (in the above figure, the qubits not in S are qubits 1 and 2); e.g., in the first subset, qubits 1 and 2 are both 0.

Consider the transformation $-W^{(S)}I_{\bar{0}}^{(S)}W^{(S)}$. Its effect is to partition the states into subsets such that in each subset the qubits that are not in S stay fixed. This transformation leaves the total probability in each subset the same—within each subset, an *inversion about average* transformation takes place. In Fig. 1, the set S contains qubits 3 and 4. It partitions the state into four subsets in which the qubits *not* in the set are fixed, e.g., in the first subset, qubits 1 and 2 are both 0. The transformation $-W^{(S)}I_{\bar{0}}^{(S)}W^{(S)}$ does an inversion about average separately in each of the four subsets.

VI. IMPROVED QUANTUM SEARCH ALGORITHM

The quantum search algorithm increases the amplitude in the t state through successive repetitions of selective inversion and inversion about average. The inversion about average operation increases the amplitude in the t state by an amount equal to the average amplitude over all states. The inversion about average requires three transformations—W, $I_{\bar{0}}$, and W—each of which requires $\log_2 N$ qubit operations. We show how to carry out the inversion about average transformations over a smaller subset of states, thus requiring fewer than $\log_2 N$ qubit operations.

A. Basic U operation

As mentioned earlier in Sec. III, the amplitude amplification principle requires a basic transformation U that produces a certain transition amplitude, U_{ts} from s to t. This can then be iterated as in Sec. III to amplify the amplitude in t.

Partition the $\log_2 N$ qubits used to represent the N items into sets of $\alpha\log_2(\log_2 N)$ qubits $(\alpha>1)$. Since there are

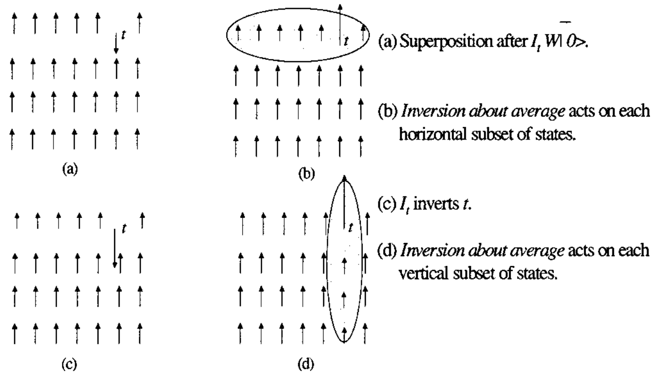
t
(a) Superposition after $I_t W|\bar{0}\rangle$.
(b) *Inversion about average* acts on each horizontal subset of states.
(c) I_t inverts t.
(d) *Inversion about average* acts on each vertical subset of states.
(a)
(b)
(c)
(d)

FIG. 2. The *inversion about average* transformation in the standard quantum search algorithm is replaced by two such operations, one that acts on the horizontal sets and the other on the vertical sets.

$\log_2 N$ qubits, there will be $\eta \equiv \log_2 N/[\alpha \log_2(\log_2 N)]$ sets (the $\eta=2$ case is depicted in Fig. 2). Define the Walsh-Hadamard transformation on the ith set as the operation $H = 1/\sqrt{2}\left[\begin{smallmatrix}1 & 1\\ 1 & -1\end{smallmatrix}\right]$, applied to each qubit in the set, and denote this by $W^{(i)}$. Similarly, define the operation $I_0^{(i)}$ as the selective inversion of the state in which each qubit in the ith set is 0. Consider the following transformation:

$$U \equiv (-W^{(\eta)} I_0^{(\eta)} W^{(\eta)}) I_t \cdots (-W^{(i)} I_0^{(i)} W^{(i)}) \times I_t \cdots (-W^{(1)} I_0^{(1)} W^{(1)}) I_t W.$$

When applied to the $|\bar{0}\rangle$ state, U has the following effect.

(i) $W|\bar{0}\rangle$ produces a superposition with equal amplitudes in all states. After this, each application of $-W^{(i)} I_0^{(i)} W^{(i)}\, I_t$ does the following.

(ii) I_t inverts the amplitude in the target state.

(iii) $-W^{(i)} I_0^{(i)} W^{(i)}$ does a partial *inversion about average* in each subset of states defined by the condition that the state of all qubits *not* in the ith set stays constant (as shown in Fig. 1).

Next consider the effect of steps (ii) and (iii) on the subset of states that contains t. Let the amplitude of the t state be $a/\sqrt{N}$. After step (ii), the amplitude of t becomes $-a/\sqrt{N}$; the amplitude of each of the other states in the subset containing t is the same as after step (i), i.e., $1/\sqrt{N}$. This is because the first $(i-1)$ inversion about average transformations acts on subsets of states in which the value of the ith qubit is constant. Hence they produce no change in the amplitude of any state in which the value of the ith qubit is different from the value of the ith qubit in the t state.

The number of states in each subset is $2^{\alpha \log_2(\log_2 N)}$, which is $(\log_2 N)^\alpha$. Therefore the average amplitude in the ith subset of states containing t is

$$\frac{1}{\sqrt{N}} - \frac{a+1}{(\log_2 N)^\alpha \sqrt{N}}.$$

Step (iii) (the partial inversion about average) increases the amplitude in t to

$$\frac{a}{\sqrt{N}} + 2\left(\frac{1}{\sqrt{N}} + \frac{a+1}{(\log_2 N)^\alpha \sqrt{N}}\right).$$

Assuming $a+1 < \log_2 N$, the increase in amplitude of t due to (ii) and (iii) is at least

$$2\left(\frac{1}{\sqrt{N}} - \frac{1}{(\log_2 N)^{\alpha-1}\sqrt{N}}\right).$$

Therefore in the η repetitions of (ii) and (iii) the amplitude of t increases by at least

$$2\eta\left(\frac{1}{\sqrt{N}} - \frac{1}{(\log_2 N)^{\alpha-1}\sqrt{N}}\right).$$

The operation U described by (i), (ii), and (iii) above forms the building block for the amplitude amplification algorithm described in the following section.

B. Amplitude amplification

As described in the analysis above, the composite operation U when applied to $|\bar{0}\rangle$ produces an amplitude of at least

$$\frac{1}{\sqrt{N}}\left[2\eta\left(1 - \frac{1}{(\log_2 N)^{\alpha-1}}\right) + 1\right]$$

in t. Therefore, by the amplitude amplification principle,

$$\frac{\pi\sqrt{N}}{4} \frac{1}{2\eta\left(1 - \frac{1}{(\log_2 N)^{\alpha-1}}\right) + 1}$$

repetitions of the $I_s U^\dagger I_t U$ operation sequence followed by a single application of U will concentrate the amplitude in the t state.

Note that $U^\dagger$ consists of the same operations as U but in the opposite order,

$$U^\dagger \equiv W I_t (-W^{(1)} I_0^{(1)} W^{(1)}) \cdots I_t (-W^{(i)} I_0^{(i)} W^{(i)}) \times \cdots I_t (-W^{(\eta)} I_0^{(\eta)} W^{(\eta)}).$$

C. Analysis

Each application of U requires η queries. Therefore, in each application of $I_s U^\dagger I_t U$ there are $(2\eta+1)$ queries. Neglecting the single application of U at the end, it follows that the total number of queries is

$$(2\eta+1) \times \frac{\pi\sqrt{N}}{4} \frac{1}{2\eta\left(1 - \frac{1}{(\log_2 N)^{\alpha-1}}\right) + 1},$$

which is less than

$$\frac{\pi\sqrt{N}}{4}\frac{1}{\left(1-\frac{1}{(\log_2 N)^{\alpha-1}}\right)}.$$

The total number of applications of U in the algorithm is

$$2\times\frac{\pi\sqrt{N}}{4}\frac{1}{2\eta\left(1-\frac{1}{(\log_2 N)^{\alpha-1}}\right)+1}$$

(as before, neglecting the single application of U at the end). The number of additional (nonquery) qubit operations required in each application of U is $\log_2 N+3\times\eta\times\alpha\log_2(\log_2 N)$, which is equal to $4\log_2 N$. The total number of additional (nonquery) qubit operations due to the U and $U^{\dagger}$ hence becomes

$$\frac{2\pi\sqrt{N}\log_2 N}{2\eta\left(1-\frac{1}{(\log_2 N)^{\alpha-1}}\right)+1}.$$

In addition there are

$$\frac{\pi\sqrt{N}}{4}\frac{1}{2\eta\left(1-\frac{1}{(\log_2 N)^{\alpha-1}}\right)+1}I_s$$

operations each of which requires $\log_2 N$ operations. Therefore, the total number of additional (nonquery) qubit operations required is

$$\frac{2\pi\sqrt{N}\log_2 N}{2\eta\left(1-\frac{1}{(\log_2 N)^{\alpha-1}}\right)+1}\times\frac{9}{8}.$$

This is less than $\frac{9}{8}\pi\alpha\sqrt{N}\log_2(\log_2 N)$ provided $\alpha\geqslant 2$.

VII. COMPARISON

The quantum search algorithm needs $\pi\sqrt{N}/4$ queries and $(3\pi\sqrt{N}\log_2 N)/4$ additional (nonquery) qubit operations. The algorithm of the previous section needs fewer than

$$\frac{\pi\sqrt{N}}{4}\frac{1}{\left(1-\frac{1}{(\log_2 N)^{\alpha-1}}\right)}$$

queries and fewer than

$$\tfrac{9}{8}\pi\alpha\sqrt{N}\log_2(\log_2 N)=(9\pi\sqrt{N}\log_2 N)/8\eta$$

additional (nonquery) qubit operations (provided $\alpha\geqslant 2$). Note that the ratio of the additional (nonquery) qubit operations required by the two algorithms is $3/2\eta$.

A. Smallest increase in the number of queries

In case $(\alpha-1)$ is $\log_2 N/2\log_2(\log_2 N)$, then the number of queries required by the improved algorithm is less than

$$\frac{\pi\sqrt{N}}{4}\frac{1}{\left(1-\frac{1}{\sqrt{N}}\right)},$$

i.e., the increase in the number of queries as compared to that required by the standard quantum search algorithm seems to be less than one. However, this is only suggestive since several other effects become significant when α becomes this large (and therefore η, the number of sets of qubits, which was $\log_2 N/[\alpha\log_2(\log_2 N)]$, becomes small). In fact, the smallest value for η is 2. We analyze this case separately below.

This is perhaps the simplest example of the partial inversion about average. The qubits are partitioned into two sets with $\frac{1}{2}\log N$ qubits in each set. Then the basic U operation is the following:

$$U\equiv(-W^{(2)}I_{\bar{0}}^{(2)}W^{(2)})I_t(-W^{(1)}I_{\bar{0}}^{(1)}W^{(1)})I_tW.$$

A simple analysis shows that the amplitude in the t state after applying U to the 0 state (which is U_{ts}) becomes $(5/\sqrt{N})-(12/N)+O(1/N^{1.5})$. An amplitude amplification as described previously in this paper will now amplify this amplitude.

To compare this to the standard quantum search algorithm, observe that the standard quantum search algorithm is obtained by taking U to be as follows:

$$U\equiv(-WI_{\bar{0}}W)I_t(-WI_{\bar{0}}W)I_tW.$$

This produces a U_{ts} of $5/\sqrt{N}+O(1/N^{1.5})$. Since the number of queries is known to be proportional to U_{ts}, the number of additional queries required by the new algorithm is obtained by scaling the queries required by the standard quantum search. This gives the number of additional queries as approximately $(\pi\sqrt{N}/4)\times(12/5\sqrt{N})\simeq 2$. Note that such a small increase in the number of queries is not likely to be significant since it would typically take the quantum search algorithm $\pi\sqrt{N}/4\pm O(1)$ queries to go from an approximate to the exact solution.

The number of additional (nonquery) qubit operations required can be compared to the standard quantum search by comparing the two U operations. Assuming each W and $I_{\bar{0}}$ need twice the number of operations as compared to $W^{(1)},W^{(2)},I_{\bar{0}}^{(1)},I_{\bar{0}}^{(2)}$, it follows that the new algorithm will need only $\frac{3}{5}$ as many operations as compared to standard quantum searching [1,8].

B. Minimizing the total number of operations

If we permit a very slight increase in the number of queries, the number of additional unitary operations and hence the total number of operations can be significantly reduced.

Assume that each query requires $K \log_2 N$ qubit operations, where K is order 1. This is plausible since the query is a function of $\log_2 N$ qubits and thus would need $O(\log_2 N)$ steps to evaluate. The total number of qubit operations is hence approximately

$$K \log_2 N \frac{\pi\sqrt{N}}{4} \frac{1}{\left(1-\frac{1}{(\log_2 N)^{\alpha-1}}\right)} + \frac{9}{8}\pi\alpha\sqrt{N}\log_2(\log_2 N)$$

$$\approx K \log_2 N \frac{\pi\sqrt{N}}{4}\left(1+\frac{1}{(\log_2 N)^{\alpha-1}}\right)$$

$$+\frac{9}{8}\pi\alpha\sqrt{N}\log_2(\log_2 N).$$

Differentiating with respect to α and setting the derivative to zero gives the condition

$$-K \log_2 N \frac{\pi\sqrt{N}}{4} \frac{\log_e(\log_2 N)}{(\log_2 N)^{\alpha-1}} + \frac{9}{8}\pi\sqrt{N}\log_2(\log_2 N) = 0.$$

This gives $(\log_2 N)^{\alpha-2} = (K \log_e 2)/4$. Substituting in the expression for the total number of operations gives

$$\frac{\pi\sqrt{N}}{4} K \log_2 N + \frac{\pi\sqrt{N}}{\log_e 2} + \frac{9}{8}\pi\sqrt{N}\log_2(\log_2 N)$$

$$+\pi\sqrt{N}\log_2 \frac{K \log_e 2}{4}$$

$$\approx \frac{\pi\sqrt{N}}{4} K \log_2 N + \frac{9}{8}\pi\sqrt{N}\log_2(\log_2 N).$$

In comparison, the standard quantum search algorithm requires

$$\frac{\pi\sqrt{N}}{4} K \log_2 N + \frac{3\pi\sqrt{N}\log_2 N}{4}$$

qubit operations. Therefore the number of additional two-qubit operations has been reduced by a factor of $\log_2 N/[3 \log_2(\log_2 N)]$.

VIII. CONCLUSION

There have been several extensions of the quantum search algorithm. This is the first improvement of the quantum search algorithm for the original exhaustive search problem. In addition, there have been several applications of the algorithm to problems not immediately related to searching (e.g., optical interferometry). Even for quantum searching, there have been implementations that relied on the unitary nature of quantum information and the phase inversion could be carried out in a single step. It is hoped that future research will extend the framework of this paper to these situations.

ACKNOWLEDGMENTS

This research was partly supported by NSA and ARO under Contract No. DAAG55-98-C-0040.

[1] L. K. Grover, Phys. Rev. Lett. **79**, 325 (1997); also at http://www.bell-labs.com/user/lkgrover/.

[2] C. H. Bennett, E. Bernstein, G. Brassard, and U. Vazirani, SIAM J. Comput. **26**, 1510 (1997).

[3] C. Zalka, Phys. Rev. A **60**, 2746 (1999).

[4] Michael L. Fredman and Dan E. Willard, J. Comput. Syst. Sci. **47**, 424 (1993).

[5] L. K. Grover, Phys. Rev. Lett. **80**, 4329 (1998).

[6] G. Brassard, P. Hoyer, and A. Tapp, e-print quant-ph/9805082; this work was originally presented in G. Brassard and P. Hoyer, e-print quant-ph/9805082.

[7] L. K. Grover, Am. J. Phys. **70**, 558 (2001); also at http://www.bell-labs.com/user/lkgrover/.

[8] Michel Boyer, Gilles Brassard, Peter Hoyer, and Alain Tapp, Fortschr. Phys. **46**, 493 (1998).