Quantum Information Review
An International Journal

http://dx.doi.org/10.18576/qir/060101

A Quantum Algorithm for Finding Common Matches between Databases with Reliable Behavior

Khaled El-Wazan

Department of Mathematics and Computer Science, Faculty of Science, Alexandria University, Egypt

Received: 2 Sep. 2017, Revised: 26 Nov. 2017, Accepted: 30 Nov. 2017
Published online: 1 Jan. 2018

Abstract: Given κ databases of unstructured entries, we propose a quantum algorithm to find the common entries between those databases. The proposed algorithm requires $\mathcal{O}(\kappa\sqrt{N})$ queries to find the common entries, where N is the number of records for each database. The proposed algorithm constructs an oracle to mark common entries, and then uses a variation of amplitude amplification technique with reliable behavior to increase the success probability of finding them.

Keywords: Quantum Algorithms, Quantum Entanglement, Databases Operations, Quantum Amplitude Amplification

1 Introduction

Given κ databases with unstructured entries, it is required to find the joint entries between those databases. Considering this problem in classical computers, an intuitive approach is to count similar entries from those databases and store them in a memory which keeps track of each entry and its number of occurrences, and then iterate over this memory and observe when the number of occurrences of certain entries equal to κ. This procedure requires at most $\mathcal{O}(\kappa N)$ steps.

Quantum computers [1,2,3] are inherently probabilistic devices which promise to significantly accelerate certain types of computations compared to classical computers [4], by utilizing quantum phenomena like entanglement and superposition. Many quantum algorithms have emerged recently, for example, Deutch-Jousza algorithm [5] that tests whether a given Boolean function is a balanced Boolean function or a constant Boolean function, using only a single oracle call. P. Shor introduced a polynomial-time algorithm [6] to factorize a composite integer to its prime factors. L. Grover presented a quantum algorithm [7] to search for an entry in an unstructured list of entries in quadratic speed-up compared to classical computers.

In 1998, Burhman *et al.* introduced an algorithm [8] that solves a problem similar to the common entries problem: given two remotely separated schedules of unknown free slots out of $N = 2^n$ slots, find a common slot between those two schedules, in as minimum communication bits sent as possible. Burhman *et al.* algorithm requires $\mathcal{O}(\sqrt{N}\log_2 N)$ communication complexity and $\mathcal{O}(k\sqrt{N})$ query calls, with k trails and error at most 2^{-k} [8]. Later in 2002, L. Grover proposed an algorithm [9] to solve scheduling problem with $\mathcal{O}(\sqrt{\varepsilon N}\log_2 N)$ computation complexity.

In 2012, Tulsi proposed a quantum algorithm [10] to find a single common element between two sets in $\mathcal{O}(\sqrt{N})$ using an ancilla qubit to mark the common solution with phase-shift and applying amplitude amplification algorithm to increase the success probability of the desired result.

In 2013, Pang *et al.* introduced a quantum algorithm [11] for set operations. In that paper, Pang *et al.* provided a subroutine to find common intersected elements between two sets of size 2^n and 2^m elements in $\mathcal{O}(\sqrt{2^{m+n}/C})$, where C is the number of common entries, using a similar algorithm proposed in [12].

The aim of this paper is to propose a quantum algorithm to find the common matches between given κ databases each of N entries. Each given database uses a black-box to identify its elements. The proposed algorithm can find a match among the common entries using a new oracle $U_{\bar{h}}$ which is constructed from the set of all given black-boxes $\bar{h}$. The new oracle $U_{\bar{h}}$ is then used along with amplitude amplification technique based on partial diffusion operator, to increase the success

* Corresponding author e-mail: khaled_elwazan@alex-sci.edu.eg

probability of finding the desired results. As well, the algorithm works with probability of success at least $2/3$.

The paper is organized as follows: Section 2 depicts an amplitude amplification algorithm with reliable behavior used to solve the problem at hand. Section 3 covers the construction of the oracle $U_{\bar{h}}$. Section 4 introduces the proposed algorithm. Section 5 analyzes the proposed algorithm. Section 6 compares the proposed algorithm to relevant literature, followed by a conclusion in Section 7.

2 Amplitude amplification

Consider having a list L of $N = 2^n$ of unstructured entries, which has an oracle U_f that is being used to access those entries. Each entry $i \in L = \{0,1,...,N-1\}$ in the list L is mapped to either 0 or 1 according to any certain property satisfied by i in L, *i.e.* $f : L \to \{0,1\}$. The unstructured search problem is stated as follows: find the entry $i \in L$ such that $f(i) = 1$.

In 1996, Grover proposed a unique approach to solve this typical problem with quadratic speed-up compared to classical algorithms [7]. The algorithm Grover proposed takes advantage of quantum parallelism to solve this problem by preparing a perfect superposition of all the possible N entries corresponding to the list L, after that it starts marking the solution using phase shift of -1 using the oracle U_f, followed by amplifying the amplitude of the solution using inversion about the mean operator. It was shown in [7,13] that the algorithm requires $\pi/4\sqrt{N}$ iteration to optimally [14] find a solution to the search problem with high probability, assuming there is only one solution $i \in L$ that satisfies the oracle U_f.

Boyer *et al.* later generalized Grover's quantum search algorithm to fit the purpose of finding multiple solutions M to the oracle U_f, *i.e.* $\forall p$, for which $1 \le p \le M \le 3N/4$, $f(i_p) = 1$, to require a number of $\pi/4\sqrt{N/M}$ iterations of the algorithm [12]. For the case of unknown number of solutions M to the oracle, an algorithm [15] was proposed to find such number M. However, the generalized quantum search algorithm has shown to exponentially fail in the case of $M > 3N/4$ [12, 13].

Younes *et al.* introduced a variation of the generalized quantum search algorithm [16] with reliable behavior in case of multiple solutions to the oracle U_f, *i.e.* $1 \le M \le N$, and requires $\mathcal{O}(\sqrt{N/M})$ oracle calls.

In the case of known multiple solutions M for a list L of size $N = 2^n$, Younes *et al.* algorithm is outlined as follows:

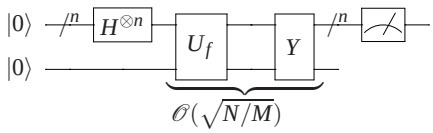

Fig. 1: Quantum circuit for the quantum search algorithm [16].

1.Prepare a quantum register with $n+1$ qubits in a uniform superposition:

$$|\varphi\rangle = \frac{1}{\sqrt{N}} \sum_{i=0}^{N-1} |i\rangle \otimes |0\rangle. \quad (1)$$

2.Iterate the algorithm for $\pi/(2\sqrt{2})\sqrt{N/M}$ times by applying the partial diffusion operator Y on the state $U_f|\varphi\rangle$ in each iteration, such that it performs the inversion about the mean on a subspace of the system, where

$$Y = (H^{\otimes n} \otimes I)(2|0\rangle\langle 0| - I_{n+1})(H^{\otimes n} \otimes I). \quad (2)$$

At any iteration $q \ge 2$, the system can be described as follows [16]:

$$|\varphi^q\rangle = a_q \textstyle\sum_{i=0}^{N-1''} (|i\rangle \otimes |0\rangle) + b_q \sum_{i=0}^{N-1'} (|i\rangle \otimes |0\rangle) + c_q \sum_{i=0}^{N-1'} (|i\rangle \otimes |1\rangle), \quad (3)$$

where the amplitudes a_q, b_q and c_q are recursively defined as follows:

$$a_q = 2\langle\alpha_q\rangle - a_{q-1}, \quad b_q = 2\langle\alpha_q\rangle - c_{q-1}, \quad c_q = -b_{q-1}, \quad (4)$$

and

$$\langle\alpha_q\rangle = \Big((1 - \frac{M}{N})a_{q-1} + (\frac{M}{N})c_{q-1}\Big). \quad (5)$$

For this algorithm, the success probability is as follows [16]:

$$P_s = \big(1 - \cos(\theta)\big)\Big(\frac{\sin^2\big((q+1)\theta\big)}{\sin^2(\theta)} + \frac{\sin^2(q\theta)}{\sin^2(\theta)}\Big), \quad (6)$$

where $\cos(\theta) = 1 - M/N$, $0 < \theta \le \pi/2$, and the required number of iterations q is given by [16]:

$$q = \Big\lfloor \frac{\pi}{2\theta} \Big\rfloor \le \frac{\pi}{2\sqrt{2}} \sqrt{\frac{N}{M}}, \quad (7)$$

where $\lfloor\ \rfloor$ is the floor operation.

Although Younes *et al.* variation of quantum search algorithm runs slower compared to Grover's algorithm by $\sqrt{2}$ for small M/N, but Younes *et al.* algorithm is more reliable with high probability than generalized Grover search algorithm for multiple matches M [16] such that $1 \le M \le N$.

3 Constructing the oracle $U_{\bar{h}}$

In this section, the given set of oracles $\bar{h}$ will be utilized to construct the oracle $U_{\bar{h}}$ which will be used for finding the common solutions M between the oracles in the set $\bar{h}$, assuming that all the given oracles are of $N = 2^n$ unstructured entries, given that n is the number of inputs to all of the given oracles. For the sake of simplification, we will provide a simple illustration for the oracle $U_{\bar{h}}$ assuming that the size of the set $\bar{h}$ is only $\kappa = 2$ oracles, and after that we will propose the generalized form of the oracle $U_{\bar{h}}$ for multiple oracles $\kappa \geq 2$.

Definition 1.*Let's assume having a Boolean function f that maps a vector of size n to either 0 or 1, i.e. $f : \{0,1\}^n \to \{0,1\}$. An oracle U_f is defined to perform such mapping. We say that U_f is an operator on $n+t+q+1$ qubits, taking the control $0 \to n-1$ qubits and targets the qubit with the index $n+t$; this configuration will be denoted as ${}^{0\to n-1}_{n+t}U_f$. Such defined oracle can be illustrated as follows:*

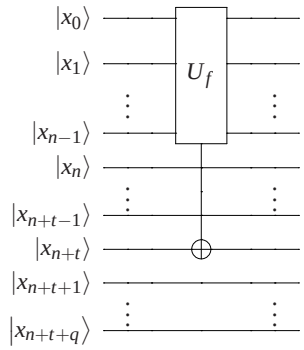

Fig. 2: A quantum circuit representing the oracle ${}^{0\to n-1}_{n+t}U_f$.

For the problem of finding common entries M between κ oracles, the problem statement can be defined as follows:

Definition 2.*Consider having a set $\mathscr{Z}$ of $\kappa \geq 2$ lists, $\mathscr{Z} = \{L_0, \cdots, L_{\kappa-1}\}$. Each list $L_j \in \mathscr{Z}$ is of $N = 2^n$ unstructured entries, which has an oracle U_j that is being used to access those entries in L_j. Each entry $i \in L_j = \{0,1,\cdots,N-1\}$ in the list L_j is mapped to either 0 or 1 according to any certain property satisfied by i in L_j, i.e. $f_j : L_j \to \{0,1\}$. The common elements problem is stated as follows: find the entry $i \in L_j$ such that $\forall L_j \in \mathscr{Z},\ f_j(i) = 1$.*

3.1 Constructing the oracle $U_{\bar{h}}$ for two databases

Given that $\kappa = 2$ oracles, U_A and U_B, which map the elements of black-box functions f_A and f_B of n input to either 0 or 1, it is required to find the common solutions M between them. It will be required to reserve 3 auxiliary qubits. An illustration of this circuit is shown in Figure 3.

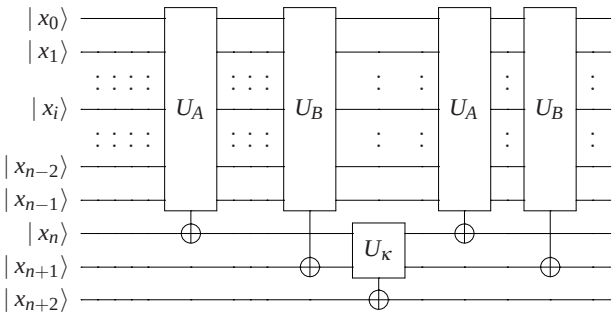

Fig. 3: A quantum circuit for the proposed oracle $U_{\bar{h}}$ for $\kappa = 2$ functions.

A quantum circuit for the oracle $U_{\bar{h}}$ can be constructed as follows:

$$U_{\bar{h}} = {}^{0\to n-1}_{n+1}U_B \times {}^{0\to n-1}_{n}U_A \times {}^{n\to n+1}_{n+2}U_\kappa \times {}^{0\to n-1}_{n+1}U_B \times {}^{0\to n-1}_{n}U_A, \tag{8}$$

where the operator U_κ represents the function $f_\kappa(x)$:

$$f_\kappa(x) = f_A(x) \cdot f_B(x), \tag{9}$$

such that $\cdot$ is the AND logic operation, and $x \in \{0,1\}^n$.

To clarify the effect of the proposed oracle $U_{\bar{h}}$, let's analyze that effect on a uniform superposition as follows:

1. *Register Preparation.* Prepare a quantum register of size $n+3$ qubits in the state $|0\rangle$, where the last 3 qubits will be utilized as extra space to compute the oracles U_A, U_B and the common solutions between them:

$$|\varphi_0\rangle = |0\rangle^{\otimes n} \otimes |0\rangle^{\otimes 3}. \tag{10}$$

2. *Register Initialization.* Apply Hadamard gates on the first n qubits to get a uniform superposition of all the possible $N = 2^n$ states:

$$\begin{aligned} |\varphi_1\rangle &= H^{\otimes n}|\varphi_0\rangle \\ &= H^{\otimes n}|0\rangle^{\otimes n} \otimes |0\rangle^{\otimes 3} \\ &= \frac{1}{\sqrt{N}} \sum_{i=0}^{N-1} |i\rangle \otimes |0\rangle^{\otimes 3}. \end{aligned} \tag{11}$$

3. *Applying the Oracle U_A.* Apply the oracle U_A on the register to mark all its possible solutions in the first extra qubit, where non-solutions will be marked with $|0\rangle$ and the solutions will be marked with $|1\rangle$:

$$|\varphi_2\rangle = {}^{0\to n-1}_{n} U_A |\varphi_1\rangle = \frac{1}{\sqrt{N}} \sum_{i=0}^{N-1} |i\rangle \otimes |f_A(i)\rangle \otimes |0\rangle^{\otimes 2}. \quad (12)$$

4. *Applying the Oracle U_B.* Apply the oracle U_B on the register to mark all its possible solutions in the second extra qubit, where the non-solution states will be marked with $|0\rangle$ and the solution states will be marked with $|1\rangle$:

$$|\varphi_3\rangle = {}^{0\to n-1}_{n+1} U_B |\varphi_2\rangle = \frac{1}{\sqrt{N}} \sum_{i=0}^{N-1} |i\rangle \otimes |f_A(i)\rangle \otimes |f_B(i)\rangle \otimes |0\rangle. \quad (13)$$

5. *Applying the Operator U_κ.* Apply the operator U_κ on the register to mark all possible common solutions between the oracles U_A and U_B in the third extra qubit, where non-common solutions will be marked with $|0\rangle$ and the common solutions will be marked with $|1\rangle$:

$$|\varphi_4\rangle = {}^{n\to n+1}_{n+2} U_\kappa |\varphi_3\rangle = \frac{1}{\sqrt{N}} \sum_{i=0}^{N-1} |i\rangle \otimes |f_A(i)\rangle \otimes |f_B(i)\rangle \otimes |f_\kappa(i)\rangle, \quad (14)$$

where $f_\kappa(i)$ is defined as in Equation (9).

6. *Applying $U_B U_A$.* Apply both the oracles $U_B U_A$ to remove any entanglement between the solutions of both oracles from the first and the second extra qubits, and reset them to their initial state $|0\rangle^{\otimes 2}$:

$$|\varphi_5\rangle = {}^{0\to n-1}_{n+1} U_B \times {}^{0\to n-1}_{n} U_A |\varphi_4\rangle = \frac{1}{\sqrt{N}} \sum_{i=0}^{N-1} |i\rangle \otimes |0\rangle^{\otimes 2} \otimes |f_\kappa(i)\rangle. \quad (15)$$

Ignoring the reset extra qubits, the state $|\varphi_5\rangle$ can be rewritten as follows:

$$|\varphi_5\rangle = \frac{1}{\sqrt{N}} \sum_{i=0}^{N-1}{}'' (|i\rangle \otimes |0\rangle) + \frac{1}{\sqrt{N}} \sum_{i=0}^{N-1}{}' (|i\rangle \otimes |1\rangle), \quad (16)$$

where $\sum''$ are all the possible uncommon solutions between the oracles U_A and U_B marked with $|0\rangle$, and $\sum'$ are all the possible common solutions between those oracles marked with $|1\rangle$.

The main reason behind applying each oracle for the second time on its target qubit when calling $U_{\bar{h}}$, is that the solutions of that specific oracle are still entangled with their target qubit. Discarding that qubit at the stage of amplifying the common solutions will drastically affect the desired outcome of the algorithm [17]. So to get rid of this entanglement, applying each oracle on its respective target qubit is necessary to remove such correlation and maintain a valid result.

3.2 The oracle $U_{\bar{h}}$ for more than two databases

Given that $\kappa \geq 2$ oracles of n input qubits and $\kappa + 1$ auxiliary qubits, we illustrate the circuit of the oracle $U_{\bar{h}}$ in Figure 4.

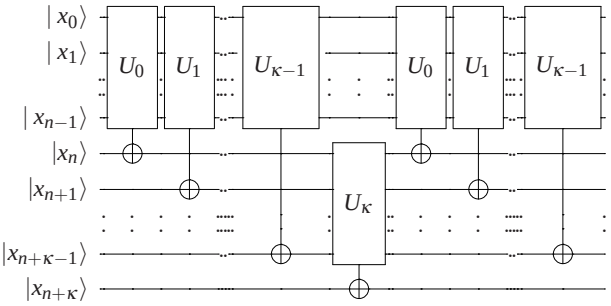

Fig. 4: A quantum circuit for the proposed oracle $U_{\bar{h}}$ for κ functions.

The oracle $U_{\bar{h}}$ can be generally defined as follows:

$$U_{\bar{h}} = \prod_{j=0}^{\kappa-1} {}^{0\to n-1}_{n+j} U_j \times {}^{n\to n+\kappa-1}_{n+\kappa} U_\kappa \times \prod_{j=0}^{\kappa-1} {}^{0\to n-1}_{n+j} U_j, \quad (17)$$

where U_κ represents the function $f_\kappa(x)$ such that

$$f_\kappa(x) = \bigwedge_{j=0}^{\kappa-1} f_j(x), \quad (18)$$

and $\bigwedge$ represents the AND logic operation.

The general system in a uniform superposition for $\kappa \geq 2$ after a single iteration, can be generally described as follows:

$$|\varphi\rangle = \frac{1}{\sqrt{N}} \sum_{i=0}^{N-1} |i\rangle \otimes |0\rangle^{\otimes \kappa} \otimes |f_\kappa(i)\rangle. \quad (19)$$

4 The proposed algorithm

In this section, we will propose the algorithm to find the common solutions M_c such that $1 \leq M_c \leq N$, among κ oracles, based on Younes *et al.* amplitude amplification algorithm. An illustration of the circuit is shown in Figure 5.

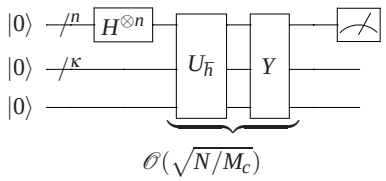

Fig. 5: Quantum circuit for the proposed algorithm.

The algorithm is carried quantum mechanically as follows:

Algorithm 1 *

1: Construct the oracle $U_{\bar{f}}$.
2: Set the quantum register to $|0\rangle^{\otimes n}$ and the extra $\kappa+1$ qubits to $|0\rangle$.
3: Apply the Hadamard gates to the first n qubits to create the uniform superposition $\frac{1}{\sqrt{N}}\sum_{i=0}^{N-1}|i\rangle\otimes|0\rangle^{\otimes\kappa+1}$.
4: Iterate over the following $q_c=\frac{\pi}{2\sqrt{2}}\sqrt{\frac{N}{M_c}}$ steps:
   1. Apply the oracle $U_{\bar{f}}$.
   2. Apply the diffusion operator Y.
5: Measure the output.

The Proposed Algorithm.

5 Analysis of the proposed algorithm

In this section, we will discuss the behavior of the proposed algorithm with respect to all possible scenarios for any given databases.

5.1 In case of known number of common matches between databases

Given that $\kappa\geq 2$ oracles, a single call to the oracle $U_{\bar{f}}$ will execute each given oracle U_j exactly 2 times. After the amplitude amplification of the desired common solutions, the total number of oracle calls q_t for all given oracles can be expressed as the following sum:

$$\begin{aligned} q_t &= \sum_{p=0}^{q_c-1} 2\times\kappa \\ &= 2\times\kappa\times\frac{\pi}{2\sqrt{2}}\sqrt{N/M_c} \\ &= \kappa\times\frac{\pi}{\sqrt{2}}\sqrt{N/M_c} \end{aligned} \tag{20}$$

So, for any given κ oracles with the same size, the number of oracle call to solve the common matches problem using the proposed algorithm is $\mathcal{O}(\kappa\sqrt{N/M_c})$.

5.2 In case of unknown matches M

An algorithm for estimating the number of matches was presented in [15], known as *quantum counting*. The proposed oracle $U_{\bar{f}}$ can be used with the quantum counting algorithm to estimate the number of matches M_c, before executing the proposed algorithm.

In [16], another algorithm was presented by Younes *et al.* to search for a match in a database, with unknown number of matches M such that $1\leq M\leq N$. Younes *et al.* algorithm can be combined with the proposed oracle $U_{\bar{f}}$ to find a common match, when the number of matches is unknown in advance.

6 Comparison with other literature

In 2012, Tulsi proposed an algorithm [10] that given two oracles that can identify the elements of two sets with the same size, the goal is to find a common element between those two sets. The success of finding that single element is further enhanced using a variation Tulsi introduced of Grover's amplitude amplification algorithm, with some restrictive conditions.

6.1 Single common solution amplification

In the case of a single common solution between $\kappa=2$ oracles, Tulsi's algorithm is found to be optimal with restrictions, and requires $\mathcal{O}(\sqrt{N})$ oracle calls. However, the proposed algorithm requires the same oracle calls $\mathcal{O}(\sqrt{N})$ but with no restrictive conditions. In the case of single common solution when $\kappa>2$ oracles which was not covered by Tulsi [10], the proposed algorithm is found to require $\mathcal{O}(\kappa\sqrt{N})$ oracle calls.

6.2 Multiple common solutions amplification

In the case of multiple common solutions between $\kappa=2$ oracles, the expected oracle calls of the proposed algorithm is $\mathcal{O}(\sqrt{N/M_c})$, when M_c is $1\leq M_c\leq N$. Tulsi's algorithm can be used to cover the case of multiple solutions when $\kappa=2$, but the problem becomes exponentially harder when $M_c>3N/4$ [12,13]. In the case of multiple common solutions between $\kappa\geq 2$ oracles, this case is not covered by Tulsi [10], however, the proposed algorithm requires $\mathcal{O}(\kappa\sqrt{N/M_c})$ oracle calls.

7 Perspective

In this paper, we proposed a quantum algorithm to find the common entries between κ databases. Each database uses an oracle to access its entries. It is shown that the

given oracles is used to construct another oracle that exhibits the behavior of finding only the common entries between those databases. The constructed oracle is used to mark the common entries with entanglement, then an amplitude amplification algorithm is applied to increase the success probability of finding the common entries.

It is found that in order to find the common matches given κ databases, it will require $\mathcal{O}(\kappa\sqrt{N/M_c})$ oracle calls. As well, It is found that the performance of the proposed algorithm is more reliable in the case of multiple matches and quadratically faster than other literature solving this problem, and handles the general case of multiple databases with similar sizes. The proposed oracle can be extended using [15] to count the number of common entries between any given oracles, or find a match as in [12,16], when the number of common entries M_c is unknown.

References

[1] S. Lloyd, *Science* **261**, 1569–1571, (1993).
[2] R. Feynman and I. Meeting. *Foundations of Physics* **16**, 507–531 (1986).
[3] D. Deutsch, *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* **400**, 97–117 (1985).
[4] E. Bernstein and U. Vazirani, *SIAM Journal on Computing* **26**, 1411–1473 (1997).
[5] D. Deutsch and R. Jozsa, *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* **439**, 553–558 (1992).
[6] P. Shor, *SIAM Journal on Computing* **26**, 1484–1509, (1997).
[7] L. Grover, *Physical Review Letters* **79**, 325 (1997).
[8] H. Buhrman, R. Cleve, and A. Wigderson, *arXiv:quant-ph/9802040*, 1-6 (1998). [Online].
[9] L. Grover, *arXiv:quant-ph/0202033*, 1–6 (2002).
[10] A. Tulsi, *arXiv:1210.4648 [quant-ph]*, 1–5 (2012).
[11] C. Pang, R. Zhou, C. Ding, and B. Hu, *Quantum Information Processing* **12**, 481–492 (2013).
[12] M. Boyer, G. Brassard, P. Høyer, and A. Tapp, *Forschritte der Physik* **46**, 493-505 (1996).
[13] A. Younes, *arXiv:0811.4481 [quant-ph]*, 1-15 (2008).
[14] C. Zalka, *Physical Review A* **60**, 2746–2751 (1999).
[15] G. Brassard, P. Høyer, and A. Tapp, *International Colloquium on Automata, Languages, and Programming*, 820-831 (1998). [Online].
[16] A. Younes, J. Rowe, and J. Miller, *Physica D: Nonlinear Phenomena* **237**, 1074–1078 (2008).
[17] M. Koashi and A. Winter, *Physical Review A - Atomic, Molecular, and Optical Physics* **69**, 6 (2004).

Khaled El-Wazan received the B.Sc. and MS.c. degree in Computer Science from Egypt. He is currently a Research and Teaching Assistant at Faculty of Science, Alexandria University, Egypt. His research interests are in the area of quantum computations and quantum algorithms. He is an active member in a research group Alexandria Quantum Computing Group AleQCG.