

# Lawrence Berkeley National Laboratory

LBL Publications

## Title

Low rank approximation in simulations of quantum algorithms

## Permalink

<https://escholarship.org/uc/item/574404tw>

## Authors

Ma, Linjian

Yang, Chao

## Publication Date

2022-03-01

## DOI

10.1016/j.jocs.2022.101561

## Copyright Information

This work is made available under the terms of a Creative Commons Attribution-NonCommercial License, available at <https://creativecommons.org/licenses/by-nc/4.0/>

Peer reviewed

# Low Rank Approximation in Simulations of Quantum Algorithms

Linjian Ma

Department of Computer Science  
University of Illinois at Urbana-Champaign  
lma16@illinois.edu

Chao Yang

Computational Research Division  
Lawrence Berkeley National Laboratory  
cyang@lbl.gov

## Abstract

Simulating quantum algorithms on classical computers is challenging when the system size, i.e., the number of qubits used in the quantum algorithm, is moderately large. However, some quantum algorithms and the corresponding quantum circuits can be simulated efficiently on a classical computer if the input quantum state is a low rank tensor and all intermediate states of the quantum algorithm can be represented or approximated by low rank tensors. In this paper, we examine the possibility of simulating a few quantum algorithms by using low-rank canonical polyadic (CP) decomposition to represent the input and all intermediate states of these algorithms. Two rank reduction algorithms are used to enable efficient simulation. We show that some of the algorithms preserve the low rank structure of the input state and can thus be efficiently simulated on a classical computer. However, the rank of the intermediate states in other quantum algorithms can increase rapidly, making efficient simulation more difficult. To some extent, such difficulty reflects the advantage or superiority of a quantum computer over a classical computer. As a result, understanding the low rank structure of a quantum algorithm allows us to identify algorithms that can benefit significantly from quantum computers.

## 1 Introduction

A quantum algorithm is often expressed by a unitary transformation  $U$  applied to a quantum state  $|\psi\rangle$ . On a quantum computer,  $|\psi\rangle$  can be efficiently encoded by  $n$  qubits, effectively representing  $2^n$  amplitudes simultaneously, and  $U$  is implemented as a sequence of one or two-qubit gates that are themselves  $2 \times 2$  or  $4 \times 4$  unitary transformations. To simulate a quantum algorithm on a classical computer, we can simply represent  $|\psi\rangle$  as a vector in  $\mathbb{C}^{2^n}$ , and  $U$  as a  $\mathbb{C}^{2^n \times 2^n}$  matrix, and perform a matrix-vector multiplication  $U|\psi\rangle$ . However, for even a moderately large  $n$ , e.g.,  $n = 50$ , the amount of memory required to store  $U$  and  $|\psi\rangle$  explicitly far exceeds what is available on many of today's powerful supercomputers, thereby making the simulation infeasible [6, 30, 31, 39]. Fortunately, for many quantum algorithms, both  $|\psi\rangle$  and  $U$  have structures. In particular,  $|\psi\rangle$  may have a low-rank tensor structure, and the quantum circuit representation of  $U$  gives a decomposition of  $U$  that can be written as

$$U = U^{(1)}U^{(2)} \dots U^{(D)}, \quad (1.1)$$

where  $U^{(i)}$  is a linear combination of Kronecker products of  $2 \times 2$  matrices, many of which are identities, and  $D$  is the depth of the circuit which is typically bounded by a (low-degree) polynomial of  $n$ . As a result, if the low rank structure of  $|\psi\rangle$  can be preserved in the successive multiplication of  $U^{(i)}$ 's with the input, we may be able to simulate the quantum algorithm efficiently for a relatively large  $n$ .

When  $|\psi\rangle$  is viewed as an **order**  $n$  tensor, there are several ways to represent it efficiently. One of them is known as a canonical polyadic (CP) decomposition [15, 17] written as

$$|\psi\rangle = \sum_{i_1, \dots, i_n \in \{0,1\}} \sum_{k=1}^R A_{i_1 k}^{(1)} A_{i_2 k}^{(2)} \dots A_{i_n k}^{(n)} |i_1 i_2 \dots i_n\rangle, \quad (1.2)$$

where  $A^{(i)} \in \mathbb{C}^{2 \times R}$  and  $R$  is known as the rank of the CP decomposition. The second representation is known as matrix product state (MPS) [34] in the physics literature or tensor train (TT) [26] in the numerical linear algebra literature, which is a special tensor networks representation of a high dimensional tensor [22]. In this representation, the quantum state can be written as

$$|\psi\rangle = \sum_{i_1, \dots, i_n \in \{0,1\}} \sum_{k_1, \dots, k_{n-1}} \mathcal{A}_{i_1 k_0 k_1}^{(1)} \mathcal{A}_{i_2 k_1 k_2}^{(2)} \cdots \mathcal{A}_{i_n k_{n-1} k_n}^{(n)} |i_1 i_2 \dots i_n\rangle, \quad (1.3)$$

where  $\mathcal{A}^{(j)}$  is a tensor of dimension  $2 \times R_{j-1} \times R_j$ , with  $R_0 = R_n = 1$ . The rank of an MPS is often defined to be the maximum of  $R_j$  for  $j \in \{1, 2, \dots, n-1\}$ . The memory requirements for CP and MPS representations of  $|\psi\rangle$  are  $\mathcal{O}(Rn)$  and  $\mathcal{O}(R^2n)$ , respectively. When  $R$  is relatively small, such requirement is much less than the  $\mathcal{O}(2^n)$  requirement for storing  $|\psi\rangle$  as a vector, which allows us to simulate a quantum algorithm with a relatively large  $n$  on a classical computer that stores and manipulates  $|\psi\rangle$  in these compact forms.

For several quantum algorithms, the rank of the CP or MPS representation of the input  $|\psi\rangle$  is low. However, when  $U^{(i)}$ 's are successively applied to  $|\psi\rangle$ , the rank of the intermediate tensors (the tensor representation of the intermediate states) can start to increase. When the rank of an intermediate tensor becomes too high, we may not be able to continue the simulation for a large  $n$ . One way to overcome this difficulty is to perform rank reductions on intermediate tensors when their ranks exceed a threshold. When a CP decomposition is used to represent  $|\psi\rangle$ , we can take, for example, (1.2) as the input and use the alternating least squares (ALS) [4, 16] algorithm to obtain an alternative CP decomposition that has a smaller  $R$ . The rank reduction of an MPS can be achieved by performing a sequence of truncated singular value decomposition (SVD).

Performing rank reduction on intermediate tensors can introduce truncation error. For some quantum algorithms, this error is zero or small, thus not affecting the final outcome of the quantum algorithm. For other algorithms, the truncation error can be large, which results in significant deviation of the computed result from the exact solution. For a specific quantum algorithm, understanding whether the intermediate tensors can be accurately approximated through low rank truncation is valuable for assessing the difficulty of simulating the algorithm on a classical computer. We attempt to investigate such difficulty for a few well known quantum algorithms in this paper both analytically and numerically.

## 1.1 Our Contributions

In this paper, we examine the use of low-rank approximation via CP decomposition to simulate several quantum algorithms. We choose to focus on using CP decomposition instead of MPS or general tensor networks to represent the input and intermediate tensors, because the (low) rank product structure of the input and intermediate tensors in the quantum algorithm are relatively easy to see and interpret in CP terms. Furthermore, some of the unitary operations such as swapping two qubits are relatively easy to implement for a CP decomposed tensor. The use of low rank MPS and more general tensor networks in quantum circuit simulation can be found in [5, 12, 14, 28, 40].

The algorithms we examine include the quantum Fourier transform (QFT) [9] and quantum phase estimation [8], which are the building blocks of other quantum algorithms, the Grover's search algorithm [2, 13], and quantum walk [7, 36] algorithms, which are quantum extensions of classical random walks on graphs.

For both QFT and phase estimation, we show that we can accurately approximate the intermediate states by low rank CP decomposition when the input states have special structures. For general input states, low-rank approximation can yield a large truncation error. For the Grover's search algorithm, we show analytically that CP ranks of all the intermediate states are bounded by  $a + 1$ , where  $a$  is the size of the marked set to be searched. Therefore, Grover's algorithm can, in principle, always be simulated efficiently by using low-rank CP decomposition when the size of the marked set is small. For quantum

walks, we show that accurate low-rank approximation is possible when the walk is performed on some graphs. However, rank reduction can be difficult when the walk is performed on a general graph.

We discuss two numerical algorithms for performing rank reduction for intermediate tensors produced in the simulation of the quantum circuit, CP-ALS and an alternative algorithm called direct elimination of scalar multiples (DESM). CP-ALS is a general and widely used algorithm for performing CP decomposition, but it may suffer from numerical issues when the initial amplitudes associated with some of the terms in CP decomposition are significantly smaller than those associated with other terms. In this case, the direct elimination of scalar multiples is more effective.

We perform numerical experiments to test the feasibility of simulating these quantum algorithms using CP decomposition. Our results show that, by using CP decomposition and low rank representation/approximation, we can indeed simulate some quantum algorithms with a many-qubit input on a classical computer with high accuracy. Other quantum algorithms such as quantum walks on a general graph are more difficult to simulate, because the CP rank of the intermediate tensors grows rapidly as we move along the depth of the quantum circuit representation of the quantum algorithm.

In summary, this paper makes the following contributions.

- We provide detailed analysis for simulating several quantum algorithms using CP decomposition.
- We discuss the feasibility of two numerical algorithms for calculating the CP decomposition in the simulation of quantum algorithms. Our results show that CP-ALS is ineffective for some algorithms, and the direct elimination of scalar multiples (DESM), can be more effective.
- We numerically show that we can accurately simulate some quantum algorithms to be run on devices consisting of 60 qubits by using CP decomposition and low rank representation/approximation.

## 1.2 Organization

This paper is organized as follows. In Section 2, we introduce the notations for quantum states, gates and circuits that are used throughout the paper. Section 3 provides the background of quantum algorithm simulations. We describe algorithms for constructing and updating low-rank CP decompositions of a tensor in 4. In Sections 5, 6 and 7, we examine the possibility of using low-rank approximations to simulate QFT and phase estimation, Grover’s search algorithm, and quantum walks, respectively. In Section 8, we compare the computational and the memory cost of simulating different quantum algorithms using CP decomposition. In Section 9, we report some numerical experimental results that demonstrate the effectiveness of using low-rank approximation to simulate quantum algorithms.

## 2 Notations for Quantum States, Gates and Circuits

Our analysis makes use of tensor algebra in both element-wise equations and specialized notation for tensor operations [19]. For vectors, lowercase Roman letters are used, e.g.,  $v$ . For matrices and quantum gates, uppercase Roman letters are used, e.g.,  $M$ . For tensors, calligraphic fonts are used, e.g.,  $\mathcal{T}$ . An order  $n$  tensor corresponds to an  $n$ -dimensional array with dimensions  $s_1 \times \dots \times s_n$ . In the following discussions, we assume that  $s_1 = \dots = s_n = 2$ . Elements of tensors are denoted in subscripts, e.g.,  $\mathcal{T}_{ijkl}$  for an order 4 tensor  $\mathcal{T}$ . For a matrix  $A$ ,  $a_i$  denotes the  $i$ th column of  $A$ . Matricization is the process of unfolding a tensor into a matrix. Given a tensor  $\mathcal{T}$  the mode- $i$  matricized version is denoted by  $T_{(i)} \in \mathbb{C}^{2 \times 2^{n-1}}$ , where all the modes except the  $i$ th mode are combined into the column. We use parenthesized superscripts to label different tensors. The Hadamard product of two matrices  $U, V$  is denoted by  $W = U * V$ . The outer product of  $n$  vectors  $u^{(1)}, \dots, u^{(n)}$  is denoted by  $\mathcal{T} = u^{(1)} \circ \dots \circ u^{(n)}$ . The Kronecker product of matrices  $A \in \mathbb{C}^{m \times n}$  and  $B \in \mathbb{C}^{p \times q}$  is denoted by  $C = A \otimes B$  where  $C \in \mathbb{C}^{mp \times nq}$ . For matrices  $A \in \mathbb{C}^{m \times k}$  and  $B \in \mathbb{C}^{n \times k}$ , their Khatri-Rao product results in a matrix of size  $mn \times k$  defined

by  $A \odot B = [a_1 \otimes b_1, \dots, a_k \otimes b_k]$ . We use  $A^\dagger$  and  $A^+$  to denote the conjugate and the pseudo-inverse of the matrix  $A$ , respectively.

The quantum state  $|\psi\rangle$  with  $n$  qubits is a unit vector in  $\mathbb{C}^{2^n}$ . It can be viewed as an order  $n$  tensor  $\mathcal{T}^{(\psi)} \in \mathbb{C}^{2 \times \dots \times 2}$ ,

$$|\psi\rangle = \sum_{i_1, \dots, i_n \in \{0,1\}} \mathcal{T}_{i_1 i_2 \dots i_n}^{(\psi)} |i_1 i_2 \dots i_n\rangle. \quad (2.1)$$

The Kronecker product of two quantum states  $|\psi\rangle, |\phi\rangle$  can be written as  $|\psi\rangle \otimes |\phi\rangle$  or  $|\psi\rangle|\phi\rangle$ . We use the quantum circuit diagram [10] to represent the unitary transformation on a  $n$ -qubit system. In the quantum circuit, the unitary transformation is decomposed into simpler unitaries according to (1.1). Each factor  $U^{(i)}$  corresponds to one layer of the circuit, which consists of Kronecker products of  $2 \times 2$  or  $4 \times 4$  unitary matrices known as one-qubit and two-qubit gates. Some commonly used one-qubit gates are:

$$H := \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}, \quad X := \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad Z := \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \quad (2.2)$$

$$R_n := \begin{bmatrix} 1 & 0 \\ 0 & e^{-\frac{2\pi i}{2^n}} \end{bmatrix}, \quad R_y(\theta) := \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}. \quad (2.3)$$

Graphically, applying  $n$   $2 \times 2$  operators  $U^{(1)}, U^{(2)}, \dots, U^{(n)}$  successively to a one-qubit state  $|x\rangle$  yields  $|y\rangle = U^{(n)} \dots U^{(2)} U^{(1)} |x\rangle$ . This operation can be drawn as

$$|x\rangle \text{ --- } \boxed{U^{(1)}} \text{ --- } \boxed{U^{(2)}} \text{ --- } \dots \text{ --- } \boxed{U^{(n)}} \text{ --- } |y\rangle.$$

The application of a  $4 \times 4$  operator  $U \otimes I$  to two qubits  $q_1$  and  $q_2$ , where  $U$  denotes an arbitrary  $2 \times 2$  unitary matrix, can be drawn as

$$\begin{array}{c} q_1 \text{ --- } \boxed{U} \text{ ---} \\ q_2 \text{ --- } \text{---} \end{array}.$$

A controlled gate controlled- $U$  is a  $4 \times 4$  operator whose expression is

$$\begin{bmatrix} I & O \\ O & U \end{bmatrix} = E_1 \otimes I + E_2 \otimes U, \quad \text{where } E_1 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \quad E_2 = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}. \quad (2.4)$$

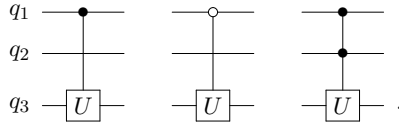
The control-on-zero gate is similar to the controlled gate and is expressed as

$$\begin{bmatrix} U & O \\ O & I \end{bmatrix} = E_1 \otimes U + E_2 \otimes I. \quad (2.5)$$

The generalized controlled gate controls the behavior of one qubit based on multiple qubits. For a 3-qubit system where  $U$  operates on the third qubit, the controlled-controlled- $U$  gate is expressed as

$$E_1 \otimes E_1 \otimes U + (I \otimes I - E_1 \otimes E_1) \otimes I = I \otimes I \otimes I + E_1 \otimes E_1 \otimes (U - I). \quad (2.6)$$

The diagrammatic representations for these three gates are shown respectively as follows,



The controlling qubit is denoted with a solid circle when it's control-on-one, and is denoted with an empty circle when it's control-on-zero. The gate  $U$  applies on the controlled qubit with a line connected to the

controlling qubits. The SWAP gate is defined as  $\text{SWAP}(|x\rangle \otimes |y\rangle) = |y\rangle \otimes |x\rangle$ , and is graphically denoted by

$$\begin{array}{c} q_1 \text{ --- } \times \text{ --- } q_2 \\ q_2 \text{ --- } \times \text{ --- } q_1 \end{array} .$$

In general, a unitary transformation  $U \in \mathbb{C}^{2^n \times 2^n}$  applied to an  $n$ -qubit state  $|\psi\rangle$  is denoted as

$$\begin{array}{c} q_1 \text{ --- } \\ \vdots \\ q_n \text{ --- } \end{array} \left[ \begin{array}{c} \text{---} \\ | \\ U \\ | \\ \text{---} \end{array} \right] \quad \text{or} \quad |\psi\rangle \xrightarrow{n} \boxed{U} \text{ ---} ,$$

where ‘ $/^n$ ’ indicates the state contains  $n$  qubits.

### 3 Simulation of Quantum Algorithms

Although tremendous progress has been made in the development of quantum computing hardware [1, 18], enormous engineering challenges remain in producing reliable quantum computers with a sufficient number of qubits required for solving practical problems. However, these challenges should not prevent us from developing quantum algorithms that can be deployed once reliable hardware becomes available. Our understanding of many quantum algorithms can be improved by simulating these algorithms on classical computers. Furthermore, classical simulations of quantum algorithms also provide a validation tool for testing quantum hardware on which quantum algorithms are to be executed.

Although we can in principle simulate quantum algorithms on a classical computer by constructing an unitary transformation as a matrix  $U$  and the input state  $|\psi\rangle$  as a vector explicitly, and performing  $U|\psi\rangle$  as a matrix vector multiplication, this approach quickly becomes infeasible as the number of simulated qubits increases.

In many quantum algorithms, the input to the quantum circuit  $|\psi\rangle$  has a low CP rank, i.e., we can rewrite  $|\psi\rangle$  as

$$|\psi\rangle = \sum_{j=1}^R a_j^{(1)} \otimes a_j^{(2)} \cdots \otimes a_j^{(n)}, \quad (3.1)$$

where  $R \ll n$  is an integer that is relatively small, and  $a_j^{(i)} \in \mathbb{C}^2$  is a vector of length 2. The storage requirement for keeping  $|\psi\rangle$  in a rank- $R$  CP format is  $\mathcal{O}(2nR)$ , which is significantly less than the  $\mathcal{O}(2^n)$  requirement for representing  $|\psi\rangle$  as an order  $n$  tensor or a single vector.

Because the unitary transformation encoded in a quantum algorithm is implemented by a quantum circuit that consists of a sequence of one- and two-qubit quantum gates as discussed in Section 2, the transformation can be implemented efficiently using local transformations that consist of multiplications of  $2 \times 2$  matrices with vectors of length 2, and we may be able to keep the intermediate states produced in the quantum circuit low rank also. Let us consider the case where the input state to a quantum circuit is rank-1, i.e.,

$$|\psi\rangle = a^{(1)} \otimes a^{(2)} \otimes \cdots \otimes a^{(n)}, \quad (3.2)$$

where  $a^{(i)} \in \mathbb{C}^2$ . It is easy to see that applying a one-qubit gate, such as the Hadamard gate  $H$ , or a two-qubit SWAP gate does not change the CP rank of  $|\psi\rangle$ . For example, if  $H$  is applied to the first qubit of  $|\psi\rangle$ , and the first and the last qubit are swapped, the resulting states become

$$Ha^{(1)} \otimes a^{(2)} \otimes \cdots \otimes a^{(n)}, \quad \text{and} \quad a^{(n)} \otimes a^{(2)} \otimes \cdots \otimes a^{(n-1)} \otimes a^{(1)}, \quad (3.3)$$

respectively. Both are still rank-1 tensors. Unfortunately, not all two-qubit gate can keep the intermediate output in rank-1. A commonly used two-qubit gate, the controlled- $U$  gate defined by (2.4), doubles the

CP rank when it is applied to a rank-1 tensor, as we can see from the simple algebraic expressions below:

$$\begin{aligned}
& (E_1 \otimes I \otimes \cdots \otimes I + E_2 \otimes U \otimes \cdots \otimes I) |\psi\rangle \\
&= E_1 a^{(1)} \otimes a^{(2)} \otimes \cdots \otimes a^{(n)} + E_2 a^{(1)} \otimes U a^{(2)} \otimes \cdots \otimes a^{(n)} \\
&= \alpha |0\rangle \otimes a^{(2)} \otimes \cdots \otimes a^{(n)} + \beta |1\rangle \otimes U a^{(2)} \otimes \cdots \otimes a^{(n)},
\end{aligned} \tag{3.4}$$

where  $\alpha, \beta$  are the first and second components of  $a^{(1)}$  respectively. As long as neither  $\alpha$  or  $\beta$  is zero, (3.4) is rank-2.

Successive applications of controlled-unitary gates where the controlling qubit vary can rapidly increase the CP rank of the output tensor. In the worst case, the rank of the output tensor can reach  $2^n$  after  $n$  controlled unitary gates are applied. This rapid increase in CP rank clearly diminishes the benefit of the low-rank representation. However, the output of several quantum algorithms are expected to have only a few large amplitude components, i.e., **they can be approximated by low rank tensors**. Therefore, the rapid increase in the CP rank of the intermediate tensors produced at successive stages of the quantum circuit may be due to the sub-optimal representation of the tensor. Because the CP decomposition of a tensor is not unique, it may be possible to find an alternative CP decomposition that has a lower rank.

When such a decomposition does not exist, we seek to find a low rank approximation that preserves the main feature of the quantum algorithm to be simulated and its output. **We discuss low rank approximation techniques in the next section and examine the effects of these low rank approximation techniques on several examples in sections 5 through 7 and section 9.** As we will see, for some quantum algorithms where **this low rank approximation yields relatively small intermediate state truncation errors**, the simulation output can still be accurate.

## 4 Low-rank Approximation in Quantum Algorithm Simulation

In this section, we discuss two techniques for reducing the rank of a CP decomposition of the tensor in the context of quantum algorithm simulation. Before we describe the details of these techniques, we first outline the basic procedure of using low rank approximation in the simulation of a quantum algorithm represented by a quantum circuit (1.1) in Algorithm 1. **Rank reduction techniques are used in Line 6 of the algorithm.**

---

### Algorithm 1 Quantum Algorithm Simulation with Low-rank Approximation

---

- 1: **Input:** An input state  $|\psi\rangle$  represented in CP format (3.1). A quantum circuit with  $D$  layers one- or two-qubit gates, i.e.,  $U = U^{(1)}U^{(2)} \cdots U^{(D)}$ , where  $U^{(i)}$  is a Kronecker product of one- or two-qubit unitaries with  $2 \times 2$  identities; maximum CP rank allowed  $r_{\max}$  for any intermediate state produced in the simulation.
  - 2: **Output:** Approximation to  $|\phi\rangle = U |\psi\rangle$ .
  - 3: **for**  $k \in \{1, 2, \dots, D\}$  **do**
  - 4:   Compute  $|\phi\rangle = U^{(k)} |\psi\rangle$ ;
  - 5:   **if** the number of rank-1 components of  $|\phi\rangle$  ( $R$  in (3.1)) exceeds  $r_{\max}$  **then**
  - 6:     Apply a rank reduction procedure to  $|\phi\rangle$  to reduce the CP rank of  $|\phi\rangle$  to at most  $r_{\max}$ ;
  - 7:   **end if**
  - 8:    $|\psi\rangle \leftarrow |\phi\rangle$ ;
  - 9: **end for**
  - 10: **return**  $|\phi\rangle$  in CP decomposed form.
- 

We should note that for some quantum algorithms, the unitary transformation  $U$  can be decomposed as

$$U = \sum_{r=1}^{R_u} A_r^{(1)} \otimes \cdots \otimes A_r^{(n)}, \tag{4.1}$$

where  $U \in \mathbb{C}^{2^n \times 2^n}$ ,  $R_u \ll 2^n$ ,  $A_r^{(i)} \in \mathbb{C}^{2 \times 2}$ ,  $i \in \{1, \dots, n\}$ . In this case, the multiplication of  $U$  with  $|\psi\rangle$  results in a low-rank tensor if  $|\psi\rangle$  is low-rank also. It is sometimes possible to obtain a good low-rank approximation of  $U$  even when  $U$  is not strictly low-rank [38]. Although seeking a low rank approximation of  $U$  can enable efficient simulations of quantum algorithms with any low-rank input states, it is a harder problem to solve than finding a low-rank approximation of intermediate tensor. In this paper, we will not discuss this approach.

To simplify our discussion, we define the matrix

$$A^{(i)} = \begin{bmatrix} a_1^{(i)} & a_2^{(i)} & \dots & a_R^{(i)} \end{bmatrix},$$

for  $i = 1, 2, \dots, n$ , where  $a_i^{(j)}$ 's are  $2 \times 1$  vectors that appear in (3.1), and sometimes use the short-hand notation

$$\{A^{(1)}, \dots, A^{(n)}\} \quad (4.2)$$

to denote the tensor  $|\psi\rangle$  (3.1) in CP representation. The  $p$ th term of (3.1) is denoted by

$$|\psi_p\rangle = \{a_p^{(1)}, a_p^{(2)}, \dots, a_p^{(n)}\}.$$

#### 4.1 Low-rank Approximation via Alternating Least Squares

We now discuss methods for performing rank reduction in Line 6 of Algorithm 1. A general way to reduce the rank of a tensor in CP format is to formulate the rank reduction problem as an optimization problem and solve the problem using a numerical optimization technique. To reduce the rank of  $|\psi\rangle$  denoted by (4.2), from  $R$  to  $s < R$ , we seek the solution to the following nonlinear least squares problem

$$\min_{B^{(1)}, \dots, B^{(n)}} \frac{1}{2} \left\| \{B^{(1)}, \dots, B^{(n)}\} - \{A^{(1)}, \dots, A^{(n)}\} \right\|_F^2, \quad (4.3)$$

where the rank of  $B^{(i)}$  is  $s$ .

Note that the target CP low-rank decomposition scenario has several differences from the standard CP decomposition scenario, where the goal is to approximate a given large tensor with low order (3 or 4). We summarize the detailed differences as follows.

- We consider the case where the input tensor order is high, and each tensor mode size is small (equals 2). The CP rank can be much higher compared to the mode size.
- We consider the case where both inputs and outputs can have complex numbers.
- In our case, the input to each CP decomposition routine is an implicit tensor expressed in the CP format with rank higher than the decomposition rank, rather than an explicit large tensor.

A widely used method for solving (4.3) is the alternating least squares (ALS) [4, 16] method, which we will refer to as the CP-ALS method. Given a starting guess of  $\{B^{(1)}, \dots, B^{(n)}\}$ , CP-ALS seeks to update one component  $B^{(i)}$  at a time while  $B^{(j)}$ 's are fixed for  $j \neq i$ . Such an update can be obtained by solving a linear least squares problem. The solution of the linear least squares problem satisfies the normal equation

$$B^{(i)} \Gamma^{(i)} = T_{(i)}^{(\psi)} P^{(i)}, \quad (4.4)$$

where  $T_{(i)}^{(\psi)} \in \mathbb{C}^{2 \times 2^{n-1}}$  is  $\mathcal{T}^{(\psi)}$  (the tensor view of  $|\psi\rangle$ ) matricized along the  $i$ th mode, the matrix  $P^{(i)} \in \mathbb{C}^{2^{n-1} \times r}$  is formed by Khatri-Rao products of  $B^{(j)}$ 's for  $j \neq i$ , i.e.,

$$P^{(i)} = B^{(1)\dagger} \odot \dots \odot B^{(i-1)\dagger} \odot B^{(i+1)\dagger} \odot \dots \odot B^{(n)\dagger}, \quad (4.5)$$



and  $\Gamma^{(i)} \in \mathbb{C}^{R \times R}$  can be computed via a sequence of Hadamard products,

$$\Gamma^{(i)} = S^{(1)} * \dots * S^{(i-1)} * S^{(i+1)} * \dots * S^{(n)}, \quad (4.6)$$

with each  $S^{(i)} = B^{(i)T} B^{(i)\dagger}$ . The *Matricized Tensor Times Khatri-Rao Product* or MTTKRP computation  $M^{(i)} = T_{(i)}^{(\psi)} P^{(i)}$  is the main computational bottleneck of CP-ALS. The computational cost of MTTKRP is  $\Theta(2^n R)$ . Because  $\mathcal{T}^{(\psi)}$  is already decomposed in CP format (4.2), the MTTKRP computation used in (4.4) can be computed efficiently via

$$M^{(i)} = A^{(i)} \left( \underset{j \in \{1, \dots, n\}, j \neq i}{*} (A^{(j)T} B^{(j)\dagger}) \right), \quad (4.7)$$

with complexity  $\mathcal{O}(nsR)$ . The algorithm is described in Algorithm 2. Consider the case where  $t$  iterations are performed in the ALS procedure, Algorithm 2 has the memory cost of  $\mathcal{O}(Rn + s^2)$  and the computational cost of  $\mathcal{O}(t(Rsn^2 + s^3n))$  (the term  $Rsn^2$  is the cost of (4.6),(4.7) and the term  $s^3n$  is the cost of performing linear system solves), yielding an asymptotic computational cost of  $\mathcal{O}(Rsn^2 + s^3n)$  considering  $t$  is usually a constant.

---

**Algorithm 2** ALS procedure for CP decomposition of an implicit tensor

---

- 1: **Input:**  $\{A^{(1)}, \dots, A^{(n)}\}$ , compression rank  $s$
  - 2: Initialize  $B^{(i)}$  as uniform random matrices within  $[0, 1]$ ,  $S^{(i)} \leftarrow B^{(i)T} B^{(i)\dagger}$  for  $i \in \{1, \dots, n\}$
  - 3: **while** not converge **do**
  - 4:   **for**  $i \in \{1, \dots, n\}$  **do**
  - 5:      $\Gamma^{(i)} \leftarrow S^{(1)} * \dots * S^{(i-1)} * S^{(i+1)} * \dots * S^{(n)}$
  - 6:     Update  $M^{(i)}$  based on (4.7)
  - 7:      $B^{(i)} \leftarrow M^{(i)} \Gamma^{(i)+}$ ,  $S^{(i)} \leftarrow B^{(i)T} B^{(i)\dagger}$
  - 8:   **end for**
  - 9: **end while**
  - 10: **return**  $\{B^{(1)}, \dots, B^{(n)}\}$
- 

## 4.2 Direct Elimination of Scalar Multiples

If the  $p$ th term in (3.1) is a scalar multiple of the  $q$ th term, for  $p \neq q$ , these two terms can be combined. As a result, the effective rank of  $|\psi\rangle$  can be lowered. As we will see in subsequent sections, scalar multiples of the same rank-1 tensor do appear in intermediate states of a quantum circuit for some quantum algorithms. Therefore, detecting such scalar multiples and combining them is an effective strategy for reducing the CP rank of intermediate states in Line 6 of Algorithm 1.

One way to check whether the  $p$ th term in (3.1) is a scalar multiple of the  $q$ th term is to compute the cosine of the angle between these two rank-1 tensors defined by

$$\cos(\theta_{p,q}) = \frac{\langle \psi_p | \psi_q \rangle}{\|\psi_p\| \cdot \|\psi_q\|}, \quad (4.8)$$

where the inner product  $\langle \psi_p | \psi_q \rangle$  can be easily computed as

$$\langle \psi_p | \psi_q \rangle = \langle a_p^{(1)}, a_q^{(1)} \rangle \cdot \langle a_p^{(2)}, a_q^{(2)} \rangle \cdots \langle a_p^{(n)}, a_q^{(n)} \rangle,$$

and  $\|\psi_p\|$  is the 2-norm of  $|\psi_p\rangle$  defined as

$$\|\psi_p\| = \sqrt{\langle \psi_p | \psi_p \rangle}.$$

If  $|\cos(\theta_{p,q})|$  is 1.0,  $|\psi_q\rangle$  is a scalar multiple of  $|\psi_p\rangle$ . It can be combined with  $|\psi_p\rangle$  as

$$|\psi_p\rangle \leftarrow \left(1 + \frac{\cos(\theta_{p,q})\beta}{\alpha}\right) |\psi_p\rangle, \quad (4.9)$$

where  $\alpha = \|\psi_p\|$  and  $\beta = \|\psi_q\|$ .

Algorithm 3 gives a procedure of detecting and combining scalar multiples of rank-1 terms in a tensor  $|\psi\rangle$  in CP format. Note that Algorithm 3 essentially computes the Gram matrix  $G$  associated with all rank-1 terms in the CP decomposition of  $|\psi\rangle$ , where the  $(p, q)$ th element of  $G$  is the cosine of the angle between the  $i$ th and  $j$ th terms. If  $G$  is rank deficient, which can be determined by performing singular value decomposition of  $G$ ,  $|\psi\rangle$  can be expressed as a linear combination of fewer tensors (viewed as vectors). However, each one of these tensor may not have a rank-1 CP form. Therefore, this approach does not necessarily yield a rank reduction in CP format.

---

**Algorithm 3** Detect and Combine rank-1 terms in a tensor  $|\psi\rangle$  in a CP format

---

```

1: Input:  $\{A^{(1)}, \dots, A^{(n)}\}$ , where  $A^{(i)} \in \mathbb{C}^{2 \times R}$ ,
2: Output:  $\{B^{(1)}, \dots, B^{(n)}\} = \{A^{(1)}, \dots, A^{(n)}\}$ , where  $B^{(i)} \in \mathbb{C}^{2 \times s}$ , with  $s \leq R$ .
3: Initialize  $B^{(i)} \leftarrow A^{(i)}$  for  $i \in \{1, \dots, n\}$ 
4:  $K \leftarrow R$ 
5:  $p \leftarrow 1$ 
6: while  $p \leq K$  do
7:    $l \leftarrow \{\}$ 
8:   for  $q \in \{p+1, \dots, K\}$  do
9:     Calculate  $\cos(\theta_{p,q})$  based on (4.8)
10:    if  $|\cos(\theta_{p,q})| = 1$  then
11:      Update  $|\psi_p\rangle := \{b_p^{(1)}, b_p^{(2)}, \dots, b_p^{(n)}\}$  based on (4.9)
12:      Append  $q$  to  $l$ 
13:    end if
14:  end for
15:  Remove the columns  $b_i^{(k)}$  from  $B^{(k)}$  when indices  $i$  appear in  $l$ , for  $k \in \{1, \dots, n\}$ 
16:   $K \leftarrow$  number of columns of  $B^{(1)}$ 
17:   $p \leftarrow p+1$ 
18: end while
19: return  $\{B^{(1)}, \dots, B^{(n)}\}$ 

```

---

### 4.3 Fidelity Estimation

Consider two states  $|\psi_P\rangle$  and  $|\psi_T\rangle$ , where  $|\psi_P\rangle$  denotes the perfect/accurate state and  $|\psi_T\rangle$  denotes the truncated (low-rank approximated) state. The fidelity  $\mathcal{F}$  of  $|\psi_T\rangle$  in approximating  $|\psi_P\rangle$  is defined as

$$\mathcal{F}(|\psi_P\rangle, |\psi_T\rangle) = |\langle \psi_P | \psi_T \rangle|^2, \quad (4.10)$$

which is a metric for measuring the accuracy of the truncated state,  $|\psi_T\rangle$ . For a general quantum algorithm, measuring the true fidelity of a low-rank approximation is generally difficult, since  $|\psi_P\rangle$  is generally not available or costly to calculate. We introduce a fidelity estimation scheme below to approximate the fidelity with much lower computational cost. This estimation scheme is used in our experiments.

Consider a circuit consisting of  $D$  layers of quantum gates. Each layer is denoted by  $U^{(i)}$ , where  $i \in \{1, \dots, D\}$ . Let the truncated state resulting from the application of the first  $i$  layers of gates be  $|\psi_T(i)\rangle$ , i.e.,  $|\psi_T(i)\rangle$  is the output of performing rank reduction on the state  $U^{(i)}|\psi_T(i-1)\rangle$ . Define the local fidelity  $f_i$  as the fidelity of this rank reduction:

$$f_i = |\langle \psi_T(i) | U^{(i)} | \psi_T(i-1) \rangle|^2, \quad (4.11)$$

the global fidelity  $\mathcal{F}$  can be approximated by the products of all the local fidelity:

$$\mathcal{F}(|\psi_P(D)\rangle, |\psi_T(D)\rangle) = |\langle \psi_P(D) | \psi_T(D) \rangle|^2 \approx \prod_{i=1}^D f_i. \quad (4.12)$$

Note that this approximation is not restricted to a specific low-rank approximation format. Our experimental results show that this approximation is accurate when approximating the state with the CP representation. Reference [40] showed that it's accurate when performing the approximation with the MPS representation.

To check the convergence of CP-ALS, we calculate the local fidelity of the CP decomposition after each ALS iteration. Consider that  $U^{(i)}|\psi_T(i-1)\rangle$  is represented in the CP format by  $\{A^{(1)}, \dots, A^{(n)}\}$  and  $|\psi_T(i)\rangle$  is represented in the CP format by  $\{B^{(1)}, \dots, B^{(n)}\}$ , the fidelity  $f_i$  can be efficiently calculated by

$$f_i = \left[ e^T \left( \underset{j \in \{1, \dots, n\}}{\ast} (B^{(j)T} A^{(j)\dagger}) \right) e \right]^2, \quad (4.13)$$

where  $e$  is an all ones vector. In this way, the computational cost of both the inner product and the fidelity calculation is  $\mathcal{O}(Rsn)$ . In addition, (4.13) can also be used to calculate the Frobenius-norm of a tensor  $\mathcal{T}$  represented in the CP representation,  $\|\mathcal{T}\|_F = \sqrt{[\langle \mathcal{T} | \mathcal{T} \rangle]}$ .

## 5 Quantum Fourier Transform and Phase Estimation

### 5.1 Quantum Fourier Transform

The quantum Fourier transform (QFT) uses a special decomposition [3,9] of the discrete Fourier transform  $F^{(N)}$  define by

$$F^{(N)} := \frac{1}{\sqrt{N}} \begin{bmatrix} \omega_N^0 & \omega_N^0 & \omega_N^0 & \cdots & \omega_N^0 \\ \omega_N^0 & \omega_N^1 & \omega_N^2 & \cdots & \omega_N^{N-1} \\ \omega_N^0 & \omega_N^2 & \omega_N^4 & \cdots & \omega_N^{2(N-1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \omega_N^0 & \omega_N^{N-1} & \omega_N^{2(N-1)} & \cdots & \omega_N^{(N-1)(N-1)} \end{bmatrix} \in \mathbb{C}^{N \times N}, \quad (5.1)$$

where  $N = 2^n$ , and the output is  $y = F^{(N)}x$  for the input vector  $x \in \mathbb{C}^N$ . We show the quantum circuit for QFT in Figure 1. As is shown in the figure, a  $n$ -qubit QFT circuit consists of  $n$  1-qubit Hadamard gates,  $\lfloor N/2 \rfloor$  SWAP gates, and  $n - 1$  controlled unitary ( $R_i$ ) gates. Without rank reduction, applying each controlled- $R_i$  gate can double the rank of the input CP tensor as shown in (3.4). The successive application of all controlled- $R_i$  gates in a QFT circuit can ultimately increase the CP rank of the output tensor exponentially.

However, for some specific input states, it is possible to represent or approximate the output tensors at different layers of the circuit with low-rank tensors. In the following theorem we show that, if the input state is a standard basis, all the intermediate states in the QFT circuit are rank 1 tensors. By a standard basis, we mean a unit vector of the form

$$|i_1\rangle \otimes |i_2\rangle \otimes \cdots \otimes |i_n\rangle, \quad (5.2)$$

where  $i_j \in \{0, 1\}$  for  $j \in \{1, 2, \dots, n\}$ .

**Theorem 5.1.** *All the intermediate states in a QFT circuit are rank 1 if the input to the circuit is a standard basis.*

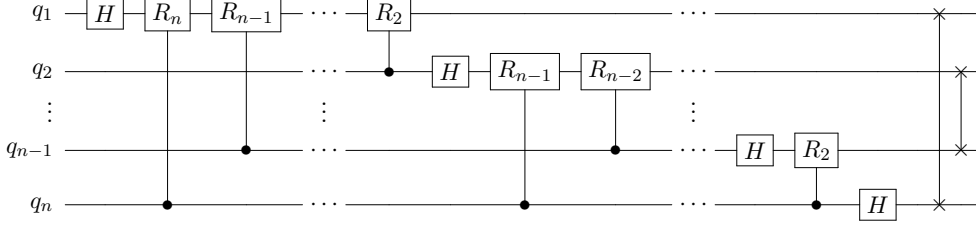


Figure 1: Circuit representation for quantum Fourier transform.

*Proof.* The proof relies on the observation that the input factor to each controlling qubit in the QFT circuit is always either  $|0\rangle$  or  $|1\rangle$ , if the input to the circuit is a standard basis. A controlled unitary does not change that factor and keeps the output as a rank-1 tensor. For example, when the first factor of the input rank-1 tensor is  $|0\rangle$  or  $|1\rangle$ , only one of the two terms in (3.4) is retained, and the rank of the output tensor remains rank-1. In addition, because the 1-qubit Hadamard ( $H$ ) gate and SWAP gate do not change the CP rank of an input tensor, all the intermediate output tensors at each layer of the circuit remain rank 1.  $\square$

Theorem 5.1 suggests that the simulation of QFT with an standard basis as the input can be simulated with  $\mathcal{O}(n)$  memory. Because the application of each 1-qubit and 2-qubit gate costs  $\mathcal{O}(1)$  operations, the overall computational cost of the simulation is  $\mathcal{O}(n^2)$ . When the input state is a linear combination of  $l$  standard basis, the CP ranks of all intermediate states are bounded by  $l$ . The memory cost for simulating such a QFT is  $\mathcal{O}(ln)$  and the computational cost of the simulation is  $\mathcal{O}(ln^2)$ .

Note that the analysis above can be extended to the analysis for the inverse of QFT ( $\text{QFT}^{-1}$ ), which inverts the input and the output of the QFT circuit shown in Figure 1. If the output of  $\text{QFT}^{-1}$  is a standard basis, then all the intermediate states will have rank 1. This is because the  $\text{QFT}^{-1}$  circuit can be expressed as

$$U = U^{(D)-1}U^{(2)-1} \dots U^{(1)-1}, \quad (5.3)$$

where  $U^{(1)}U^{(2)} \dots U^{(D)}$  makes the QFT circuit. Therefore, the intermediates of  $\text{QFT}^{-1}$  are the same as those in QFT, but in a reversed order.

## 5.2 Phase Estimation

One of the main applications of QFT is phase estimation [8]. The goal of phase estimation is to estimate an eigenvalue of a unitary operator  $U$  corresponding to a specific eigenvector  $|\psi\rangle$ . All eigenvalues of  $U$  are on the unit circle, which can be represented by  $e^{i2\pi\theta}$  for some phase angle  $\theta$ . We assume that  $|\psi\rangle$  can be prepared somehow, and there exists an “oracle” that performs  $U^{2^j}|\psi\rangle$  for  $j \in \{0, \dots, n-1\}$ .

The quantum circuit that performs phase estimation takes two registers as the input. The first register is initialized as  $|0 \dots 0\rangle$ . The number of qubits ( $n$ ) contained in the register depends on how accurately we want to represent  $\theta$  as a binary. The second register is used to prepare the target eigenvector  $|\psi\rangle$ . The circuit consists of a set of Hadamard gates applied to the first register, followed by a sequence of controlled- $U^{2^j}$  gates as shown in Figure 2. After applying these gates, we obtain the following rank-1 tensor in the first register,

$$\frac{1}{2^{n/2}} \left( |0\rangle + e^{i2\pi 2^{n-1}\theta} |1\rangle \right) \otimes \dots \otimes \left( |0\rangle + e^{i2\pi 2^1\theta} |1\rangle \right) \otimes \left( |0\rangle + e^{i2\pi 2^0\theta} |1\rangle \right). \quad (5.4)$$

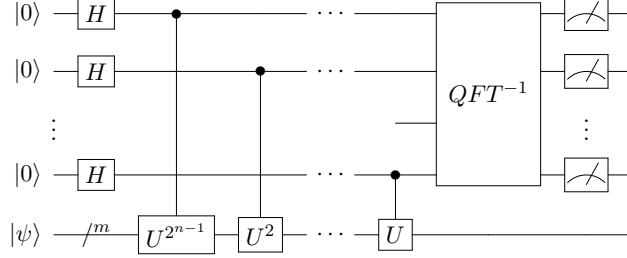


Figure 2: Circuit representation for phase estimation.

When the above state is used as the input to  $QFT^{-1}$ , we obtain the following output,

$$\frac{1}{2^n} \sum_{x=0}^{2^n-1} \sum_{k=0}^{2^n-1} e^{-\frac{2\pi ik}{2^n}(x-2^n\theta)} |x\rangle. \quad (5.5)$$

If  $\theta$  satisfies

$$x = 2^n\theta, \quad (5.6)$$

the amplitude of  $|x\rangle$  is 1 and the amplitude of  $|y\rangle$  is zero for  $y \neq x$ . As a result, (5.5) is a standard basis and is rank-1. Furthermore, because the inverse QFT circuit is identical to the QFT circuit, but with the input and output reversed, the intermediate output at each layer of the inverse QFT circuit should be rank-1 when (5.4) is the input and (5.6) holds exactly. In this case, the phase estimation algorithm can be simulated efficiently.

When (5.6) holds only approximately, rank reduction will be needed during the simulation of the inverse QFT to keep the intermediate output at each layer of the inverse QFT circuit low rank. We can use the techniques discussed in section 4 to perform rank reduction. In Appendix A we provide an analysis that shows all the intermediate states on the first register of the phase estimation circuit can be approximated by a low-rank state whose CP rank is bounded by  $\mathcal{O}(1/\epsilon)$ , and the output state fidelity is at least  $1 - \epsilon$ .

When the input state of the second register is a linear combination of  $l$  eigenvectors, for  $l \ll n$ , which occurs in applications such as noisy phase estimation [27] and NMRS quantum walk based search [32], the output of the circuit has  $l$  large amplitudes. This type of phase estimation can also be simulated efficiently.

## 6 Grover's Algorithm

Search is a common problem in information science. Grover's algorithm [13] achieves quadratic speed-up compared to the classical search algorithms. We first examine the possibility to simulate the Grover's algorithm with only one marked item using the CP representation of the tensor, then generalize the analysis to the cases with multiple marked items.

### 6.1 Search with One Marked Item

The goal of the search problem is to find a particular item  $x^*$  called the marked item from a set  $(X)$  of  $N = 2^n$  items that contains  $x^*$ . On a classical computer, the worst case complexity of finding  $x^*$  is  $\mathcal{O}(N)$ . On a quantum computer, one can use the Grover's algorithm to find  $x^*$  in  $\mathcal{O}(\sqrt{N})$  steps. In this

algorithm, each item in the set to be searched is mapped to a basis state in an  $2^n$ -dimensional Hilbert space. The algorithm involves applying an unitary transformation of the form,

$$U^{(g)} = U^{(o)}U^{(d)}, \quad (6.1)$$

successively to a superposition of all basis states,  $|h\rangle = H^{\otimes n}|0^n\rangle$ , where  $U^{(o)}$  is known as an *oracle* that recognizes the item to be searched but does not provide the location of the item, and  $U^{(d)}$  is known as a diffusion operator, which is a reflector to be defined below. It is well known that after  $\mathcal{O}(\sqrt{N})$  successive applications of  $U^{(g)}$  to  $|h\rangle$ , the amplitude associated with  $|x^*\rangle$  becomes close to 1, while the amplitudes associated with other basis states become close to 0.

The oracle can be defined as

$$U^{(o)}|x\rangle = (-1)^{f(x)}|x\rangle, \quad \text{where } f(x) = \begin{cases} 1 & \text{if } x = x^*, \\ 0 & \text{otherwise.} \end{cases} \quad (6.2)$$

This oracle is a unitary, which can be implemented as a generalized controlled-NOT gate show in Figure 3. The diffusion operator is defined as

$$U^{(d)} = 2|h\rangle\langle h| - I. \quad (6.3)$$

Because  $|h\rangle$  can be obtained by applying Kronecker products of Hadamard matrices to the standard basis state  $|0^n\rangle$ , we can write (6.3) as

$$U^{(d)} = H^{\otimes n} (2|0\rangle\langle 0|^{\otimes n} - I) H^{\otimes n}. \quad (6.4)$$

This unitary can be implemented by a layer of Hadamard gates followed by a generalized controlled-NOT, followed by another layer of Hadamard gates as shown in Figure 3.

We show below that all the intermediate states produced at different layers of the circuit for Grover's algorithm can be represented by a linear combination of  $|h\rangle$  and  $|x^*\rangle$ , which are both rank-1, if the input to the circuit is  $|h\rangle$ .

Without loss of generality, let us assume that  $U^{(g)}$  is applied to the input  $|x\rangle = \alpha|h\rangle + \beta|x^*\rangle$ . When  $|x\rangle$  is the input to the entire circuit, we have  $\alpha = 1$  and  $\beta = 0$ . Applying  $U^{(o)}$  to  $|x\rangle$  yields

$$U^{(o)}|x\rangle = \alpha \left( |h\rangle - \frac{1}{\sqrt{N}}|x^*\rangle \right) - \alpha \frac{1}{\sqrt{N}}|x^*\rangle - \beta|x^*\rangle = \alpha|h\rangle - \left( \alpha \frac{2}{\sqrt{N}} + \beta \right) |x^*\rangle, \quad (6.5)$$

which remains in the span of  $|h\rangle$  and  $|x^*\rangle$ . The application of  $U_d$  to  $U_o|x\rangle$  starts with the application of  $H^{\otimes n}$  to last expression in (6.5). This yields  $\alpha|0^n\rangle - (\alpha \frac{2}{\sqrt{N}} + \beta)H^{\otimes n}|x^*\rangle$ , which is in the span of  $|0^n\rangle$  and  $H^{\otimes n}|x^*\rangle$ . The subsequent application of the reflector  $2|0^n\rangle\langle 0^n| - I$  still keeps the result in  $\text{span}\{|0\rangle, H^{\otimes n}|x^*\rangle\}$ . Applying  $H^{\otimes n}$  again to  $|0\rangle$  and  $H^{\otimes n}|x^*\rangle$  respectively brings them back to  $|h\rangle$  and  $|x^*\rangle$ . Therefore, the CP ranks of all intermediate states are at most 2.

## 6.2 Search with Multiple Marked Items

We now generalize the analysis to the cases with multiple marked items. The problem can be cast as the problem of finding an arbitrary element in the non-empty marked subset,  $A$ , from the set  $X$ . When mapped to the quantum circuit, the set  $A$  and the unmarked set  $B$  can be expressed as

$$\begin{aligned} A &= \{x \in \{0, 1\}^n : f(x) = 1\}, \\ B &= \{x \in \{0, 1\}^n : f(x) = 0\}. \end{aligned}$$

Let  $a = |A|$  and  $b = |B|$ , we have  $a + b = N = 2^n$ . Grover's algorithm starts from  $|h\rangle$ , and applies the operator  $U^{(g)}$  for  $\lfloor \frac{\pi}{4} \sqrt{\frac{N}{a}} \rfloor$  times. Let

$$|A\rangle = \frac{1}{\sqrt{a}} \sum_{x \in A} |x\rangle \quad \text{and} \quad |B\rangle = \frac{1}{\sqrt{b}} \sum_{x \in B} |x\rangle, \quad (6.6)$$

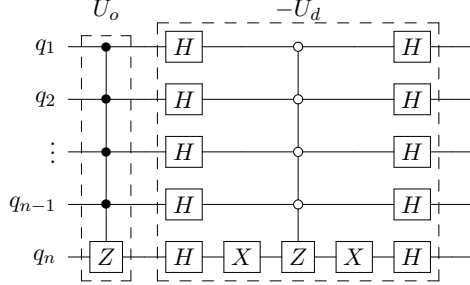


Figure 3: Circuit implementation for the operator  $U^{(g)}$  in Grover's algorithm. The implementation assumes that the marked state is  $|x^*\rangle = |11\dots 1\rangle$ .

as is shown in [23], application of  $U^{(g)}$  on  $|A\rangle$  and  $|B\rangle$  results in

$$U^{(g)}|A\rangle = \left(1 - \frac{2a}{N}\right)|A\rangle - \frac{2\sqrt{ab}}{N}|B\rangle \quad \text{and} \quad U^{(g)}|B\rangle = \frac{2\sqrt{ab}}{N}|A\rangle - \left(1 - \frac{2b}{N}\right)|B\rangle. \quad (6.7)$$

Because the initial state is in the space  $S$  spanned by  $|A\rangle$  and  $|B\rangle$ , all the intermediate states are all in the space  $S$ .

It is clear that  $|A\rangle$  is low rank if  $a \ll 2^n$  because it can be written as a linear combination of  $a$  standard bases. Although  $|B\rangle$  may not appear to be low rank because it is a linear combination of many standard bases, it is actually low rank because  $|B\rangle$  can be written as  $|B\rangle = (2^{n/2}|h\rangle - \sqrt{a}|A\rangle)/\sqrt{b}$ . Since  $|h\rangle$  is rank-1, the rank of  $|B\rangle$  is  $a + 1$ . As a result, all intermediate states in the Grover's algorithm are low rank.

The observation we made above suggests that the states that emerge from the applications of  $U^{(g)}$  are low-rank. However, since both  $U^{(o)}$  and  $U^{(d)}$  are implemented with controlled-unitary gates, applying  $U^{(g)}$  directly will increase the rank of the intermediate states. Rank reduction techniques described in Section 4 need to be used to keep the ranks of these tensors low. The complexity of the rank reduction procedure is thus closely related to the rank of the intermediate states emerging from direct applications of  $U^{(g)}$  to the input. A detailed analysis, which is presented in Appendix B, shows that the rank of the state input to the rank reduction procedure is at most  $2(a + 1)$  times the optimal rank  $(a + 1)$ , making the procedure still efficient because  $a$  is small. For example, consider a 5-qubit system and  $a = 2$ . The two marked items are  $|11111\rangle$  and  $|00000\rangle$ . Then

$$\begin{aligned} U^{(o)} &= (I - 2E_1 \otimes E_1 \otimes E_1 \otimes E_1 \otimes E_1)(I - 2E_0 \otimes E_0 \otimes E_0 \otimes E_0 \otimes E_0) \\ &= I - 2E_1 \otimes E_1 \otimes E_1 \otimes E_1 \otimes E_1 - 2E_0 \otimes E_0 \otimes E_0 \otimes E_0 \otimes E_0. \end{aligned}$$

If the input state to  $U^{(o)}$  has rank  $R$ , the output from applying  $U^{(o)}$  to the input has rank  $(a + 1)R = 3R$ . The subsequent application of the gate  $U^{(d)}$  increases the rank by at most a factor of two. We outline the simulation of Grover's algorithm using rank reduction in Algorithm 4.

We can use either a direct elimination of scalar multiples (DESM) or CP-ALS to reduce the rank of an intermediate tensor. For Grover's algorithm, DESM is much more efficient. This is because the output that emerge from the direct application of  $U^{(g)}$  contains several terms that are scalar multiples of each other. For the example system shown above, applying  $U^{(o)}$  to a state that's in the space  $S$  yields

$$U^{(o)}(\alpha|00000\rangle + \beta|11111\rangle + \gamma|h\rangle)$$

$$= \alpha(|00000\rangle - 2|00000\rangle) + \beta(|11111\rangle - 2|11111\rangle) + \gamma(|h\rangle - \frac{1}{2\sqrt{2}}|00000\rangle - \frac{1}{2\sqrt{2}}|11111\rangle).$$

We can see that applying DESM can reduce the rank of the output tensor to 3. A similar reduction can be achieved after  $U^{(d)}$  is applied.

---

**Algorithm 4** Simulating Grover's algorithm with rank reduction

---

- 1: **Input:** Input state  $|h\rangle$  represented in CP format  $\{A^{(1)}, \dots, A^{(n)}\}$ , maximum rank allowed  $r_{\max}$
  - 2: **for** iter = 1, 2, ...,  $\lfloor \frac{\pi}{4} \sqrt{N} \rfloor$  **do**
  - 3:    $\{B^{(1)}, \dots, B^{(n)}\} \leftarrow$  Apply  $U^{(g)}$  to the state represented by  $\{A^{(1)}, \dots, A^{(n)}\}$
  - 4:    $\{A^{(1)}, \dots, A^{(n)}\} \leftarrow$  Rank.Reduction( $\{B^{(1)}, \dots, B^{(n)}\}, r_{\max}$ )
  - 5:   Normalize the state represented by  $\{A^{(1)}, \dots, A^{(n)}\}$
  - 6: **end for**
  - 7: **return**  $\{A^{(1)}, \dots, A^{(n)}\}$
- 

The analysis above assumes that  $a$  is known in advance. In cases where  $a$  is unknown, it may not be easy to set the rank threshold. In Appendix B, we also show that for any marked set that is small in size, we can approximate the intermediate states by rank-2 states, and the simulation can still yield one marked state with high probability after  $\mathcal{O}(\sqrt{N})$  steps of Grover's algorithm.

## 7 Quantum Walks

Quantum walks [21, 37] play an important role in the development of many quantum algorithms, including quantum search algorithms [35] and the quantum page rank algorithm [29]. A quantum walk operator is the quantum extension of a classical random walk operator that has been studied extensively in several scientific disciplines. A classical random walk is characterized by an  $N \times N$  Markov chain stochastic matrix  $P$  associated a graph with  $N$  vertices. There is an edge from the  $j$ th vertex to the  $i$ th vertex if the  $(i, j)$ th element of  $P$ , denoted by  $P_{ij}$ , is nonzero. A random walk is a process in which a walker randomly moves from vertex  $j$  to vertex  $i$  with probability  $P_{ij}$ . If  $v$  is a vector that gives a probability distribution of the initial position of the walker, then  $w = P^t v$  gives the probability distribution of the walker's position after  $t$  steps of the random walk are taken. One of the key results in the classical random walk theory is that the standard deviation of the walker's position with respect to its mean position after taking  $t$  steps is  $\mathcal{O}(\sqrt{t})$ . In contrast, the standard deviation is known to be  $\mathcal{O}(t)$  in a corresponding quantum walk.

The simplest type of coined quantum walk on a one dimensional cyclic lattice can be represented by the following unitary matrix

$$U = (I \otimes H)(L \otimes E_1 + R \otimes E_2), \quad (7.1)$$

where  $H, E_1$  and  $E_2$  are defined in (2.2), (2.4), and  $L$  and  $R$  are left and right shift permutation matrices defined by

$$L = \begin{pmatrix} 0 & 0 & \dots & 0 & 1 \\ 1 & 0 & \ddots & \ddots & 0 \\ 0 & 1 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \dots & \dots & 1 & 0 \end{pmatrix}, \quad R = \begin{pmatrix} 0 & 1 & \dots & \dots & 0 \\ 0 & 0 & 1 & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 1 \\ 1 & 0 & \dots & \ddots & 0 \end{pmatrix}. \quad (7.2)$$

The permutation matrices  $L$  and  $R$  correspond to the stochastic matrix  $P$  in a classical random walk. The Hadamard matrix  $H$  is known as a quantum coin operator that introduces an additional degrees of freedom in determining how the walker should move on the 1D lattice.

A quantum walk on a more general graph defined by a vertex set  $V$  and edge set  $E$  can be described by a formalism established by Szegedy [32, 36]. Szegedy's quantum walk is defined on the edges of the



bipartite cover of the original graph  $(V, E)$ , i.e. the graph is mapped to a Hilbert space  $\mathcal{H}^{|V|^2} = \mathcal{H}^{|V|} \otimes \mathcal{H}^{|V|}$ , with the orthonormal computational basis defined as follows:

$$\{|x, y\rangle := |x\rangle \otimes |y\rangle : x \in V, y \in V\}. \quad (7.3)$$

For each  $x \in V$ ,  $|\psi^{(x)}\rangle$  is defined as the weighted superposition of the edges emanating from  $x$ ,

$$|\psi^{(x)}\rangle = |x\rangle \otimes \sum_{y \in V} \sqrt{P_{yx}} |y\rangle = |x\rangle \otimes |\phi^{(x)}\rangle. \quad (7.4)$$

By making use of the SWAP operator  $S$  and the reflection operator  $U^{(d)}$  associated with  $\{|\psi^{(x)}\rangle : x \in V\}$ , defined as

$$S = \sum_{x, y \in V} |y, x\rangle \langle x, y|, \quad (7.5)$$

$$U^{(d)} = 2 \sum_{x \in V} |\psi^{(x)}\rangle \langle \psi^{(x)}| - I = \sum_{x \in V} |x\rangle \langle x| \otimes \left( 2 |\phi^{(x)}\rangle \langle \phi^{(x)}| - I \right),$$

we can define a quantum walk operator as

$$U^{(w)} = SU^{(d)}. \quad (7.6)$$

A Szegedy's quantum walk can be used as a building block for searching a marked vertex in a graph. Let  $x^*$  be the vertex to be searched. The oracle associated with  $x^*$  is defined as

$$U^{(o)} = I - 2 \sum_{y \in V} |x^*, y\rangle \langle x^*, y|. \quad (7.7)$$

Using such an oracle, we can perform the search by applying the following unitary [33],

$$U^{(s)} = U^{(o)} U^{(w)} U^{(w)}, \quad (7.8)$$

to an initial state.

In the following, we will show quantum circuits for implementing  $U^{(s)}$  associated with a few Markov transition matrices induced by structured graphs. We also assume the transition probability from vertex  $x$  to vertex  $y$  is defined by

$$P_{yx} = \frac{A_{yx}}{\text{outdeg}(x)}, \quad (7.9)$$

where  $A$  is the adjacent matrix of the graph and  $\text{outdeg}(x)$  is the number of edges emanating from  $x$ . We will show that for some structured graphs, the quantum search can be simulated efficiently if the initial state is rank-1.

### 7.1 Quantum Walk on a Complete Graph with Self-loops

We first consider a complete graph with  $|V| = N = 2^n$  vertices. We assume that a random step from a vertex can return to the vertex itself, i.e., the graph contains self-loops. The transition probability (Markov chain) matrix  $P$  associated with such a random walk can be expressed as

$$P = \frac{1}{N} e e^T, \quad (7.10)$$

where  $e$  is a vector of all ones. In this case, the vector  $|\phi^{(x)}\rangle$  defined in (7.4) is the same for all  $x \in V$ , and can be set to  $|h\rangle = \frac{1}{\sqrt{N}} \sum_{x \in V} |x\rangle$ . As a result, the diffusion operator in (7.5) can be simplified to

$$U^{(d)} = 2 \sum_{x \in V} |x\rangle \langle x| \otimes |h\rangle \langle h| - I = I \otimes (2|h\rangle \langle h| - I). \quad (7.11)$$

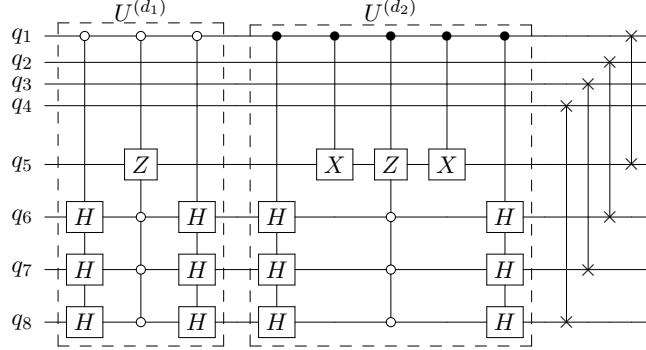


Figure 4: Circuit implementation of the operator  $U^{(d)}$  for a quantum walk on a complete bipartite graph. The implementation assumes that  $n_1 = n_2 = 3$ .

Likewise, the oracle operator in (7.7) can be simplified to

$$U^{(o)} = I - 2|x^*\rangle\langle x^*| \otimes \left( \sum_{y \in V} |y\rangle\langle y| \right) = (I - 2|x^*\rangle\langle x^*|) \otimes I. \quad (7.12)$$

With these simplified expression for  $U^{(d)}$  and  $U^{(o)}$ , the search operator defined in (7.8) has the form

$$\begin{aligned} U^{(s)} &= \left( (I - 2|x^*\rangle\langle x^*|) \otimes I \right) \left( (2|h\rangle\langle h| - I) \otimes I \right) \left( I \otimes (2|h\rangle\langle h| - I) \right) \\ &= \left( (I - 2|x^*\rangle\langle x^*|)(2|h\rangle\langle h| - I) \right) \otimes \left( 2|h\rangle\langle h| - I \right). \end{aligned} \quad (7.13)$$

The circuit implementation of  $U^{(s)}$  is similar to the implementation for the Grover's algorithm, since the operator on the first  $n$  qubits is  $U^{(g)}$  (expressed in (6.1) and shown in Figure 3), and the operator on the last  $n$  qubits is (6.3). It is easy to verify that, when the  $U^{(s)}$  is applied to the rank-1 tensor  $|h\rangle|h\rangle$ , it produces a tensor with a rank that is at most 2. As a result, the search for  $x^*$  can be simulated on a classical computer with  $\mathcal{O}(n)$  complexity in flops and memory.

## 7.2 Quantum Walk on a Complete Bipartite Graph

We now consider a quantum walk on a complete bipartite graph  $K_{N_1, N_2}$ , where the set of vertices  $V$  consists of two subsets  $V_1$  and  $V_2$ , where  $N_1 = |V_1| = 2^{n_1}$ ,  $N_2 = |V_2| = 2^{n_2}$ . Each vertex in  $V_1$  is connected to all vertices in  $V_2$  and vice versa. In this case vector  $|\phi^{(x)}\rangle$  is defined by  $|h^{(V_1)}\rangle = \frac{1}{\sqrt{N_2}} \sum_{x \in V_2} |x\rangle$ , for  $x \in V_1$ , and by  $|h^{(V_2)}\rangle = \frac{1}{\sqrt{N_1}} \sum_{x \in V_1} |x\rangle$  for  $x \in V_2$ . The diffusion operator can be rewritten as

$$U^{(d)} = \sum_{x \in V_1} |x\rangle\langle x| \otimes \left( 2|h^{(V_1)}\rangle\langle h^{(V_1)}| - I \right) + \sum_{x \in V_2} |x\rangle\langle x| \otimes \left( 2|h^{(V_2)}\rangle\langle h^{(V_2)}| - I \right). \quad (7.14)$$

If we start from a rank-1 quantum state, all intermediate quantum states in a quantum walk based search on a complete bipartite graphs can be low-rank. Below we provide analysis for  $n_1 = n_2$ , and the analysis can be generalized to the cases where  $n_1 \neq n_2$ .

We can map all  $2 \cdot 2^{n_1}$  vertices to quantum states described by  $n_1 + 1$  qubits. The quantum circuit for the corresponding quantum walk operates on  $2n + 2$  qubits. Each vertex  $x \in V_1$  can be represented by

$|0\rangle|x\rangle$ , and each vertex  $x \in V_2$  can be represented by  $|1\rangle|x\rangle$ . Using this representation and the definition  $|h\rangle \equiv H|0\rangle$ , we obtain

$$|h^{(V_1)}\rangle = |1\rangle|h\rangle^{\otimes n_1}, \quad \text{and} \quad |h^{(V_2)}\rangle = |0\rangle|h\rangle^{\otimes n_1}, \quad (7.15)$$

$$\sum_{x \in V_1} |x\rangle\langle x| = |0\rangle\langle 0| \otimes I^{\otimes n_1}, \quad \text{and} \quad \sum_{x \in V_2} |x\rangle\langle x| = |1\rangle\langle 1| \otimes I^{\otimes n_1}. \quad (7.16)$$

Assuming that  $x^* \in V_1$ , the oracle operator in (7.7) can be simplified to

$$U^{(o)} = I - 2|0\rangle\langle 0| |x^*\rangle\langle x^*| \otimes \left( \sum_{y \in V_2} |y\rangle\langle y| \right) = I - 2|0\rangle\langle 0| |x^*\rangle\langle x^*| \otimes |1\rangle\langle 1| \otimes I^{\otimes n_1}. \quad (7.17)$$

Based on (7.16),  $U^{(d)}$  can be rewritten as

$$U^{(d)} = \underbrace{|0\rangle\langle 0| \otimes I^{\otimes n_1} \otimes \left( 2|1\rangle\langle 1| \otimes |h\rangle^{\otimes n_1} \langle h|^{\otimes n_1} - I^{\otimes(n_1+1)} \right)}_{U^{(d_1)}} + \underbrace{|1\rangle\langle 1| \otimes I^{\otimes n_1} \otimes \left( 2|0\rangle\langle 0| \otimes |h\rangle^{\otimes n_1} \langle h|^{\otimes n_1} - I^{\otimes(n_1+1)} \right)}_{U^{(d_2)}}.$$

As a result, we can rewrite  $U^{(d_1)}$  and  $U^{(d_2)}$  as

$$U^{(d_1)} = |0\rangle\langle 0| \otimes I^{\otimes n_1} \otimes \left[ (I \otimes H^{\otimes n_1}) \left( 2|1\rangle\langle 1| \otimes |0\rangle^{\otimes n_1} \langle 0|^{\otimes n_1} - I^{\otimes(n_1+1)} \right) (I \otimes H^{\otimes n_1}) \right], \quad (7.18)$$

$$U^{(d_2)} = |1\rangle\langle 1| \otimes I^{\otimes n_1} \otimes \left[ (I \otimes H^{\otimes n_1}) \left( 2|0\rangle\langle 0| \otimes |0\rangle^{\otimes(n_1+1)} \langle 0|^{\otimes(n_1+1)} - I^{\otimes(n_1+1)} \right) (I \otimes H^{\otimes n_1}) \right]. \quad (7.19)$$

The circuit implementations of these operators are shown in Figure 4.

The input state to the quantum walk based search algorithm is the superposition of all edge states,

$$\frac{1}{\sqrt{2|V_1|}} \sum_{x \in V_1} |x\rangle|\phi^{(x)}\rangle + \frac{1}{\sqrt{2|V_2|}} \sum_{x \in V_2} |x\rangle|\phi^{(x)}\rangle = \frac{1}{\sqrt{2}} |0\rangle|h\rangle^{\otimes n_1} |1\rangle|h\rangle^{\otimes n_1} + \frac{1}{\sqrt{2}} |1\rangle|h\rangle^{\otimes n_1} |0\rangle|h\rangle^{\otimes n_1}.$$

Applying the search operator  $U^{(s)} = U^{(o)}U^{(w)}U^{(w)} = U^{(o)}SU^{(d)}SU^{(d)}$  amounts to applying  $U^{(d)}$ ,  $SU^{(d)}S$  and  $U^{(o)}$  successively to the input state. Similar to the analysis for Grover's algorithm, it can be verified that the intermediate states emerging from the application of these unitaries implemented in quantum circuit are in the space spanned by

$$\{|0\rangle|h\rangle^{\otimes n_1} |1\rangle|h\rangle^{\otimes n_1}, |1\rangle|h\rangle^{\otimes n_1} |0\rangle|h\rangle^{\otimes n_1}, |0\rangle|x^*\rangle |1\rangle|h\rangle^{\otimes n_1}, |1\rangle|h\rangle^{\otimes n_1} |0\rangle|x^*\rangle\}.$$

The CP rank of these intermediate states is bounded by 4. As a result, both the overall memory and computational costs are bounded by  $\mathcal{O}(n)$ .

### 7.3 Quantum Walk on Cyclic Graphs

The last type of quantum walk we examine is performed on a cyclic graph. The transition probability matrix associated with a cyclic graph is a circulant matrix. We focus on a subset of cyclic graphs in which each vertex is connected to  $N - a$  other vertices, where  $N = 2^n$  and  $a = 2^m - 1$  for some natural number  $m < n$ . The matrix elements of the transition matrix  $P$  are defined as

$$P_{yx} = \begin{cases} \frac{1}{\sqrt{N-a}} & \text{if } (y-x) \bmod N \geq a, \\ 0 & \text{otherwise.} \end{cases} \quad (7.20)$$

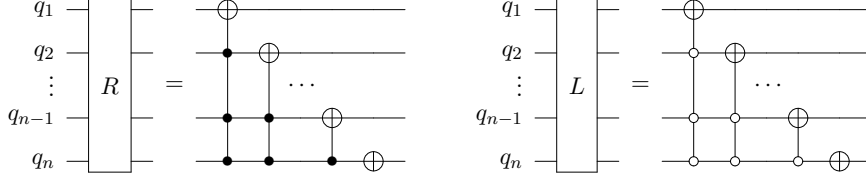


Figure 5: Circuit implementation for the one-element rotation operators [11].

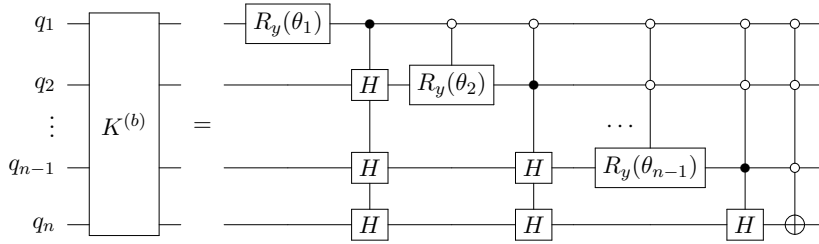


Figure 6: Circuit implementation for  $K^{(b)}$  for the complete graph without self-loops.  $\theta_i$  is defined as  $\theta_i = \arccos \sqrt{\frac{2^{n-i}-1}{2^{n-i+1}-1}}$ .

When  $a = 1$ , the cyclic graph becomes a complete graph. To simplify the notation, we use  $|x\rangle$  to denote the state  $|x \bmod N\rangle$ . We also let  $|v^{(x)}\rangle$  be the superposition of vertices that are not adjacent to  $x$ ,

$$|v^{(x)}\rangle = \frac{1}{\sqrt{a}} \sum_{i=0}^{a-1} |x+i\rangle. \quad (7.21)$$

It follows that the vector  $|\psi^{(x)}\rangle$  defined in (7.4) can be rewritten as

$$|\psi^{(x)}\rangle = |x\rangle \left( \sqrt{\frac{N}{N-a}} |h\rangle - \sqrt{\frac{a}{N-a}} |v^{(x)}\rangle \right). \quad (7.22)$$

Consequently, the diffusion operator  $U^{(d)}$  has the form

$$U^{(d)} = \left( 2 \frac{N}{N-a} I \otimes |h\rangle\langle h| - I \right) + \sum_{x \in V} \frac{2a}{N-a} |x\rangle\langle x| \otimes \left( -\sqrt{\frac{N}{a}} (|v^{(x)}\rangle\langle h| + |h\rangle\langle v^{(x)}|) + |v^{(x)}\rangle\langle v^{(x)}| \right). \quad (7.23)$$

When  $a \ll N$ ,  $U^{(d)}$  is dominated by the first term, making it behave like a diffusion operator that appears in the Grover's algorithm discussed in section 6. However, the presence of the second term can introduce entanglement in the intermediate states in a quantum circuit implementation of  $U^{(d)}$ . Our experimental results in Section 9.4 shows that low-rank approximation for quantum walk based search on cyclic graphs is not accurate.

To implement  $U^{(d)}$  as a quantum circuit, we use the technique presented in [20] to construct a unitary operator  $U^{(t)}$  that diagonalizes  $U^{(d)}$ . It has been shown that, for cyclic graphs with a circulant  $P$ , we can construct a unitary  $U^{(t)}$  so that

$$U^{(t)} U^{(d)} U^{(t)+} = I \otimes (2|b\rangle\langle b| - I), \quad (7.24)$$

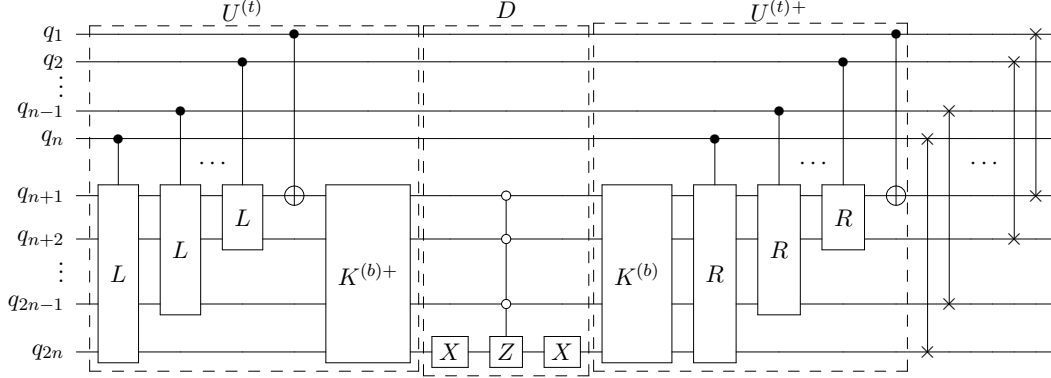


Figure 7: Circuit implementation for the quantum walk  $U^{(w)}$  on cyclic graphs.

for some computational basis  $|b\rangle$ . The diagonal unitary  $D = 2|b\rangle\langle b| - I$  can be efficiently implemented by a quantum circuit. The unitary  $U^{(t)}$  can be written as a product of shift permutation matrices  $L$  of different sizes, tensor product with identities, as well a unitary  $K^{(b)}$  that maps the first column of  $P$  to  $|b\rangle$ . Shift permutation can be implemented efficiently by circuits shown in Figure 5. When each column of  $P$  is sparse or structured,  $K^{(b)}$  can also be implemented by an efficient quantum circuit (for the complete graph without self-loops, the implementation is shown in Figure 6). As a result,  $U^{(d)} = U^{(t)+}DU^{(t)}$  can be implemented by an efficient circuit shown in Figure 7.

To summarize, quantum walk based search can be accurately low-rank simulated only on specific structure graphs, including complete graphs with self-loops, and complete bipartite graphs. For general graphs, such as cyclic graphs, the low-rank simulation will not be accurate.

## 8 Summary of Computational Cost

The use of low rank CP decomposition to represent the input and intermediate states in the simulation of a quantum algorithm allows us to significantly reduce the memory requirement of the simulation. If the rank of all intermediate states can be bounded by a small constant, then the memory requirement of the simulation is linear with respect to the number of qubits  $n$ . This is significantly less than the memory required to simulate a quantum algorithm directly, which is exponential with respect to  $n$ , by performing a matrix-vector multiplication.

By keeping the input and intermediate states in low rank CP form, we can also significantly reduce the number of floating point operations (FLOPs) in the simulation. Because a quantum gate in each layer of the quantum circuit is typically local, meaning that it is a  $2 \times 2$  matrix operating on one factor of a CP term, the number of FLOPs required to multiply quantum gates with the input states is proportional to  $nrD$ , where  $r$  is the maximum rank of all intermediate states and  $D$  is the depth of the quantum circuit. Therefore, as long as  $D$  and  $r$  are not too large, the cost of applying a unitary transformation in decomposed form (i.e. a quantum circuit) to a low rank input is relatively low also. However, to keep intermediate states in low rank CP decompositions, rank reduction through CP-ALS or **direct elimination of scalar multiples (DESM, Algorithm 3)** may need to be used. The cost of the rank reduction computation can be higher than applying the unitary transformation associated with the quantum algorithm.

We summarize the computational costs of simulating the quantum algorithms analyzed above using CP decomposition in Table 1. **To lower the rank of intermediate tensors in the simulation, the cost of**

DESM (Algorithm 3) is lower than CP-ALS (Algorithm 2). However, as can be seen in the table, DESM cannot always be effectively applied to the simulation of a quantum algorithm in which an intermediate tensor does not contain CP terms that are multiple of each other. The CP-ALS is a more general method to compress intermediate tensors emerging from successive layers of a quantum circuit. It can be used for the simulation of all quantum algorithms. When both DESM and CP-ALS can be applied, the cost of CP-ALS is typically much higher even when the number of ALS iterations is fixed at a small constant. Therefore, whenever possible, we should try using DESM first before using CP-ALS.

Table 1: The asymptotic computational cost of simulating different quantum algorithms with CP low-rank approximation. The computational cost shown for Grover’s algorithm is the cost of applying each  $U^{(g)}$  defined in (6.1), and for quantum walks is the cost of applying each  $U^{(w)}$  defined in (7.6).

Algorithm	DESM	CP-ALS
QFT w/ standard basis input state	$\mathcal{O}(n^2)$	$\mathcal{O}(n^3)$
QFT w/ random rank-1 input state and CP rank limit $r$	/	$\mathcal{O}(r^2n^3 + r^3n^2)$
Phase estimation w/ CP rank limit $r$	/	$\mathcal{O}(r^2n^3 + r^3n^2)$
Grover’s Algorithm w/ $a$ marked items	$\mathcal{O}(a^3n)$	$\mathcal{O}(a^3n^2)$
Quantum walks based search w/ complete graph with loops	$\mathcal{O}(n)$	$\mathcal{O}(n^2)$
Quantum walks based search w/ complete bipartite graph	$\mathcal{O}(n)$	$\mathcal{O}(n^2)$

## 9 Experimental Results

In this section, we demonstrate the efficacy of using low rank approximation in the simulation of four quantum algorithms: QFT, phase estimation, Grover’s algorithm and quantum walks. We implemented our algorithms on top of an open-source Python library, “Koala” [28], which is a quantum circuit/state simulator and provides interface to several numerical libraries, including NumPy [25] for CPU executions and CuPy [24] for GPU executions. All of our code is available at <https://github.com/LinjianMa/koala>. Most of our experiments were carried out on an Intel Core i7 2.9 GHz Quad-Core machine using NumPy routines. For some QFT simulation experiments with large number of qubits and large CP rank limits, the experiments were carried out on an NVIDIA Titan X GPU using CuPy routines to accelerate the execution.

In each of these simulations, the input to the simulated quantum circuit is chosen to be a rank-1 or low rank state in the CP representation. We limit the CP rank of the intermediate states to a fixed integer that varies from one algorithm to another. This limit may also change with respect to the number of simulated qubits employed in the quantum algorithm. Typically, a higher limit is required for simulations that employ more qubits.

We measure the fidelity of the simulation output using the metric defined in Section 4.3. To improve the accuracy of the CP-ALS rank reduction procedure, we repeat the procedure several times using different initial guesses of the approximate low-rank CP tensors. The approximation that yields the highest fidelity is chosen as the input to the next stage of the simulation.

### 9.1 Quantum Fourier Transform

We first simulate the QFT algorithm when the input is a standard basis. As is discussed in Theorem 5.1, all the intermediate states can be represented by a rank-1 state. The results are shown in in Table 2. As can be seen, we can accurately simulate the large circuits with as many as 60 qubits.

We also simulate the QFT algorithm in which the input to the QFT circuit is a randomly generated rank-1 state in the CP representation. Each element of each CP factor is drawn from a uniformly distributed random number between 0 and 1. As we explained in Section 5.1, even though the input state to the QFT circuit has rank 1, the output of the circuit may not be low rank. Furthermore, the

Table 2: Numerical experimental results for QFT with DESM when the input is a standard basis. The rank limit ( $r$ ) is set as 1 for all the experiments.

Number of qubits	18	20	22	24	26	28	30	32	40	60
Fidelity estimation (4.12)	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0

rank of the intermediate states resulting from the application of a sequence of one or two-qubit gates may increase rapidly if no rank reduction procedure is applied.

We simulate the QFT algorithm performed on circuits with at least 16 qubits and at most 40 qubits, with the results presented in Table 3. The experiments with 40 qubits are run on one GPU, and all the other experiments are run on a CPU. These simulations correspond to Fourier transforms of vectors of dimensions between  $2^{16} = 65,536$  and  $2^{40} \approx 1.1 \times 10^{12}$ . For all the experiments, we use CP-ALS to reduce the rank of the intermediate states when the CP rank of the state exceeds the limit  $r$  reported in the second row of Table 3. Three different random initial guesses are used in each attempt to reduce the rank of the intermediate state.

We report the approximation fidelity of the simulation output in the third row of Table 3. As can be seen from this table, fidelity beyond 90% can be achieved for 16 to 24-qubit QFT simulations when the CP rank of all intermediate states are limited to 256. As the number of qubits increase to 26, 27 or 28, the fidelity of the output drops below 90%. Higher CP ranks are necessary to maintain high fidelity. The largest system we have tested has 40 qubits. Even when we increase the limit of the CP rank to 2048, we can only achieve 58% fidelity. We didn't further increase the rank limit, since the simulation time will be too long (the experiment with rank limit being 2048 took around 10 hours to finish).

Table 3: Numerical experimental results for QFT with CP-ALS when the input is a random state with CP rank 1.

Number of qubits	16	20	24	26	27	28	40	40
Rank limit ( $r$ )	256	256	256	256	256	256	1024	2048
Fidelity estimation (4.12)	0.998	0.975	0.918	0.784	0.845	0.788	0.534	0.580

## 9.2 Phase Estimation

We next simulate phase estimation by constructing a rank-1 state according to (5.4) as the input to an inverse QFT quantum circuit. We set the phase angle  $\theta$  in (5.4) to  $\theta = 1/2(1 + 1/2^n)$ . It follows from (A.3) that this particular choice of  $\theta$  results in an output state that is not simply an elementary basis. According to Theorem A.1, the algorithm can still be efficiently low-rank approximated.

The CP rank of the intermediate state resulting from the application of a two-qubit gate doubles. We use CP-ALS to reduce the rank of the state when the rank becomes larger than the limit of 20. Three random initial guesses are used in each CP-ALS rank reduction step, and the best approximation is used to continue the simulation.

The fidelity of the simulation is shown in Table 4. Because all intermediate states can be well approximated by rank-20 states, we are able to simulate large circuits with as many as 60 qubits. As can be seen from the table, high fidelity ( $> 0.999$ ) can be achieved for all simulations.

Table 4: The fidelity of phase estimation simulation.

Number of qubits	18	20	22	24	26	28	30	32	40	60
Rank limit ( $r$ )	20	20	20	20	20	20	20	20	20	20
Fidelity estimation (4.12)	1.0	0.9997	0.9998	0.9998	0.9995	0.9993	0.9993	0.9997	0.9972	0.9994

### 9.3 Grover’s Algorithm

In Section 6, we showed that Grover’s algorithm is intrinsically low-rank, i.e., when the input to the Grover’s quantum circuit is a particular rank-1 state, all intermediate states produced at each layer of the circuit have low ranks. Therefore, we should be able to simulate the algorithm by keeping the rank of all intermediate states in the simulation low using CP-ALS.

However, in practice, the use of CP-ALS to reduce the rank of an intermediate state can be difficult for large circuits that contain many qubits. This difficulty arises from the large disparity between the magnitudes of different CP components in the intermediate state produced in the early iterations of the Grover’s algorithm. To be specific, the intermediate state produced by the first iteration of the Grover’s algorithm can be expressed as  $\alpha|h\rangle + \beta|x^*\rangle$ , where the coefficient  $\beta$  has a magnitude close to  $1/\sqrt{N}$  or  $1/\sqrt{2^n}$ , whereas  $\alpha \sim O(1)$ . The amplitude of  $\beta$  decreases exponentially with respect to the number of qubits. For a large  $n$ , we found that the CP-ALS output were in the direction of  $|h\rangle$  for most of the random initial guesses, and the amplitude of  $|x^*\rangle$  were not effectively amplified. As a result, it is difficult for CP-ALS to numerically identify the  $x^*$  component in the intermediate state.

Table 5 shows the feasibility of using CP-ALS in the simulation of Grover’s algorithm. We use Grover’s circuits that encode one marked item to be searched ( $|A| = 1$ ) as well as circuits that encode 20 marked items to be searched ( $|A| = 20$ ). For all the experiments, we set the CP rank limit to 2. In each experiment, we use multiple random initial guesses in the first iteration of the Grover’s algorithm to produce a low rank approximation. The number of initial guesses are listed in the row labelled by Num-ALS-init. The use of more initial guesses can yield a more accurate low rank approximation. In subsequent iterations of Grover’s algorithm, we use the CP-ALS approximation produced in the previous iterations as the initial guess. In exact arithmetic, this scheme guarantees that all intermediate tensors produced in the simulated Grover’s algorithm lie within the subspace spanned by  $|h\rangle$  and  $|x^*\rangle$ , and the amplitude of  $|x^*\rangle$  is amplified incrementally.

Although we can use the metric defined in (4.12) to estimate the fidelity of the low rank approximation in the Grover’s algorithm, the amplitude amplification feature of the Grover’s algorithm allows us to assess the fidelity of low rank approximation by directly measuring amplitudes (coefficients) associated with the basis of the marked items in the final output state. To be specific, for a set of marked items  $A$  represented by  $|A|$  computational basis of an  $n$ -qubit state, the fidelity of the approximation can be evaluated as

$$\sum_{x \in A} |\langle \psi | x \rangle|^2, \tag{9.1}$$

where  $|\psi\rangle$  is the simulation output. As the number of Grover’s iterations approaches  $\frac{\pi}{4}\sqrt{N}$ , the amplitude sum in (9.1) should become close to 1.0.

As we can see from Table 5, when  $|A| = 1$ , the amplitude of the marked item becomes close to 1.0 when the number of qubits is less or equal to 14 and 3 ALS initial guesses are used. When the number of qubits reaches 16, 3 ALS initial guesses are not enough for an accurate simulation, and we need to try 10 different initial guesses in the CP-ALS algorithm to obtain an accurate rank-2 approximation to the output of Grover’s circuit. When  $|A| = 20$ , the amplitude sum obtained at the end of all simulations are above 0.5. However, for a 10-qubit simulation, the amplitude sum is only slightly above 0.5. For a 14-qubit simulation, the amplitude sum is 0.6. These low amplitude sums indicate the difficulty of using CP-ALS to find an accurate low-rank approximation to intermediate states in some iterations of the Grover’s algorithm. The success of CP-ALS depends on the random initial guesses used in the first iteration of the Grover’s algorithm. For a 16 qubit simulation, we were able to obtain an amplitude sum that is close to 1.0. This is likely due to a good initial guess generated for CP-ALS.

The difficulty encountered in CP-ALS rank reduction can be easily overcome by the use of the DESM reduction technique aimed at identifying CP components that are scalar multiples of each other. This is the technique we discussed in Algorithm 3. Table 6 shows that by using this technique we can consistently



simulate Grover’s algorithm for single or multiple marked search items with high accuracy. We have simulated circuits with as many as 30 qubits. In all experiments, the amplitude sums of the marked items obtained at the end of the simulation are close to 1.0.

Table 5: Numerical experimental results for Grover’s algorithm with CP-ALS. The rank limit ( $r$ ) is set as 2 for all the experiments.

Number of qubits	8	10	12	14	16	16	8	10	12	14	16
Num-marked-item	1	1	1	1	1	1	20	20	20	20	20
Num-ALS-init	3	3	3	3	3	10	3	3	3	3	3
Amplitude of $A$ (9.1)	1.	0.999	1.0	1.0	0.0	1.0	0.972	0.537	0.981	0.607	0.998

Table 6: Numerical experimental results for Grover’s algorithm with DESM. The rank limit ( $r$ ) is set as 2 for all the experiments.

Number of qubits	10	15	20	25	30	10	15	20	25	30
Num-marked-item	1	1	1	1	1	20	20	20	20	20
Amplitude of $A$ (9.1)	0.999	1.0	1.0	1.0	1.0	0.999	1.0	1.0	1.0	1.0

#### 9.4 Quantum Walks

In Section 7, we showed that intermediate states in a quantum walk can be low-rank when the starting point of the walk is a particular rank-1 state for some graphs (e.g. complete graphs with loops and complete bipartite graphs). We will show that in these cases, the quantum walk can be efficiently and accurately simulated using a low rank representation.

We measure the accuracy of the simulations by the estimated fidelity (4.12) and the amplitude (coefficient) of the marked item,  $x^*$ , which is defined as

$$\sum_{y \in V} |\langle \psi | x^*, y \rangle|^2, \quad (9.2)$$

where  $|\psi\rangle$  is the output state. High amplitude and high fidelity mean that the approximated algorithm is accurate. Similar to Grover’s algorithm, when the number of iterations approaches  $\frac{\pi}{4}\sqrt{N}$ , the amplitude sum in (9.2) should become large (greater than 0.5 for all test cases, and close to 1 for systems with a large number of qubits).

Tables 7 and 8 show the experimental results for quantum walks on complete graphs with loops, using CP-ALS and DESM, respectively. As can be seen, we can accurately simulate the algorithm using low rank approximation by DESM for systems with different qubit counts. The rank of all intermediate tensors in these simulations are limited to 2. For CP-ALS, we find that for a large number of qubits, imposing a small rank limit (2) usually cannot guarantee accuracy in the rank-reduction procedure. For example, the amplitude of  $x^*$  for the 20-qubit system is close to 0, as is shown in Tables 7. The reason is that CP-ALS tends to fail for intermediate states with a large disparity in the magnitudes of different CP components, similar to what is discussed in Section 9.3 for Grover’s algorithm. To be specific, the intermediate state produced by the first iteration of the quantum walk algorithm can be expressed as  $\alpha|h\rangle|h\rangle + \beta|x^*\rangle|h\rangle$ , where the amplitude of  $\beta$  decreases exponentially with respect to the number of qubits. For a large  $n$ , the CP-ALS output were in the direction of  $|h\rangle|h\rangle$  for most of the random initial guesses, and the amplitude of  $|x^*\rangle$  were not effectively amplified. To achieve high accuracy, we need to slightly increase the rank limit. Similar to the behavior of the algorithm observed when the number of initial guesses for ALS is increased, increasing the rank limit of the intermediate tensor tends to improve the likelihood of finding better approximations in CP-ALS. For example, when we increase the rank limit

to 5, the algorithm can be accurately simulated for the 20-qubits' system, and when we further increase the rank limit to 20, we can accurately simulate the 24-qubits' system.

Tables 9 and 10 show the experimental results for quantum walks on complete bipartite graphs obtained from using CP-ALS and DESM rank reduction techniques, respectively. As discussed in Section 7.2, all intermediate states can be accurately represented by rank 4 tensors. As can be seen, we can accurately simulate these quantum walks with DESM for graphs of various sizes. We set the rank limit to 4 in these simulations. For CP-ALS, similar to the simulation of quantum walks on complete graphs with loops, we need to slightly increase the rank limit to achieve high accuracy. Without a sufficient increase in the allowed rank of the intermediate tensor, the CP-ALS procedure may produce a result that has a small or zero amplitude for  $x^*$ . As we can see in Table 9, the amplitude of  $x^*$  becomes 0.0 for the 20-qubit and 24-qubit runs. However, when we increase the rank limit to 40, the algorithm can be accurately simulated for a graphs with  $2^{24}$  vertices.

Table 7: Numerical experimental results for quantum walks on complete graphs with loops with CP-ALS. The number of initial guesses in CP-ALS is set as 3 for all the experiments.

Number of qubits	12	16	20	20	24	24
Rank limit $r$	2	2	2	5	5	20
Amplitude of $x^*$ (9.2)	0.964	0.983	0.002	1.0	0.0	0.999

Table 8: Numerical experimental results for quantum walks on complete graphs with loops with DESM. The rank limit ( $r$ ) is set as 2 for all the experiments. All the experiments have 1.0 fidelity.

Number of qubits	12	16	20	24	28	32
Amplitude of $x^*$ (9.2)	0.964	0.983	0.998	0.999	1.0	1.0

Table 9: Numerical experimental results for quantum walks on complete bipartite graphs with CP-ALS. The number of initial guesses in CP-ALS is set as 3 for all the experiments.

Number of qubits	8	12	16	20	20	24	24
Rank limit $r$	4	4	4	4	10	10	40
Amplitude of $x^*$ (9.2)	0.781	0.897	0.942	0.0	0.988	0.0	0.998

Table 10: Numerical experimental results for quantum walks on complete bipartite graphs with DESM. The rank limit ( $r$ ) is set as 4 for all the experiments. All the experiments have 1.0 fidelity.

Number of qubits	8	12	16	20	24	28	32	36
Amplitude of $x^*$ (9.2)	0.781	0.897	0.942	0.988	0.998	1.0	1.0	1.0

We also perform experiments to show that, for a general cyclic graph, a quantum walk based search algorithm is more difficult to simulate because the rank of the intermediate states increases rapidly and it is more difficult to reduce the rank of the intermediate states by using CP-ALS. We test on complete graphs (without loops) with one marked item  $x^*$ . The results are shown in Table 11. For all experiments, we use 3 different random CP-ALS initializations for each low-rank truncation routine. As can be seen in the table, as the number of qubits increases, the CP rank threshold necessarily needed to reach high simulation accuracy also increases exponentially. As a result, it becomes more difficult to simulate the quantum walk on larger graphs.

Although the low-rank simulation of this algorithm via CP-ALS is not accurate, CP decomposition requires lower memory cost compared to the naive state vector representation. As can be seen in Table 11, when the number of qubits is  $N$ , a CP rank of  $2^{N/2+1}$  yields accurate results, and the memory cost is only  $(2N) \cdot 2^{N/2+1}$ .

Table 11: Numerical experimental results for quantum walk on complete graphs with CP-ALS.

Number of qubits	6	10	14
Rank	$2^4$	$2^6$	$2^8$
Fidelity estimation (4.12)	1.0	0.840	0.998
Amplitude of $x^*$	0.645	0.807	0.938

## 10 Conclusions

In this paper, we examined the possibility of using low-rank approximation via CP decomposition to simulate quantum algorithms on classical computers. The quantum algorithms we have considered include the quantum Fourier transform, phase estimation, Grover’s algorithm and quantum walks.

For QFT, we have shown that all the intermediate states within the QFT quantum circuit and the output of the transform are rank-1 when the input is a standard (computational) basis. The same observation holds for the phase estimation algorithm, i.e., all the intermediate states in an phase estimation algorithm can be accurately approximated by a low-rank tensor. When the input to the QFT circuit is a general rank-1 tensor, the CP rank of the intermediate states can grow rapidly. Applying rank reduction in the simulation of the QFT can lead to loss of fidelity in the output.

For Grover’s algorithms, we have shown that the CP ranks of all the intermediate states are bounded by  $a + 1$ , where  $a$  is the size of the marked set. Therefore, Grover’s algorithm can, in principle, always be simulated efficiently by using low-rank CP decomposition when the size of the marked set is small.

For quantum walks, we have shown that the algorithm can be simulated efficiently on some graphs such as complete graphs with loops and complete bipartite graphs when the transition probability along edges of the graph is constant. We point out that it may be difficult to simulate a quantum walk defined on a more general graph, e.g., a general cyclic graph with non-uniform transition probabilities.

We presented two methods for performing rank reduction for intermediate tensors produced in the simulation of the quantum circuit. The CPD-ALS is a more general approach. However, it may suffer from numerical issues when the initial amplitudes associated with some of the terms in CP decomposition is significantly smaller than those associated with other terms. In this case, a method based on a direct elimination of scalar multiples is more effective.

Our numerical experimental results demonstrate that, by using CP decomposition and low rank representation/approximation, we can indeed simulate some quantum algorithms with a many-qubit input on a classical computer with high accuracy. Other quantum algorithms such as quantum walks on a more general graph with non-uniform transition probabilities are more difficult to simulate, **because the CP rank of the intermediate tensors grows exponentially with respect to the system size (number of qubits), and low-rank approximations cannot maintain sufficient accuracy.** This difficulty in fact demonstrates the real advantage or superiority of a quantum computer over a classical computer for solving certain classes of problems.

## Acknowledgements

This work is supported by the U.S. Department of Energy (DOE) under Contract No. DE-AC02-05CH11231, through the Office of Advanced Scientific Computing Research Accelerated Research for Quantum Computing Program and the SciDAC Program.

## References

- [1] Frank Arute, Kunal Arya, Ryan Babbush, Dave Bacon, Joseph C Bardin, Rami Barends, Rupak Biswas, Sergio Boixo, Fernando GSL Brandao, David A Buell, et al. Quantum supremacy using a programmable superconducting processor. *Nature*, 574(7779):505–510, 2019.

- [2] Gilles Brassard, Michele Mosca, and Alain Tapp. Quantum amplitude amplification and estimation. 2002.
- [3] Daan Camps, Roel Van Beeumen, and Chao Yang. Quantum Fourier transform revisited. *arXiv preprint arXiv:2003.03011*, 2020.
- [4] J Douglas Carroll and Jih-Jie Chang. Analysis of individual differences in multidimensional scaling via an N-way generalization of “Eckart-Young” decomposition. *Psychometrika*, 35(3):283–319, 1970.
- [5] Claudio Chamon and Eduardo R Mucciolo. Virtual parallel computing and a search algorithm using matrix product states. *Physical review letters*, 109(3):030503, 2012.
- [6] Zhao-Yun Chen, Qi Zhou, Cheng Xue, Xia Yang, Guang-Can Guo, and Guo-Ping Guo. 64-qubit quantum circuit simulation. *Science Bulletin*, 63(15):964–971, 2018.
- [7] Andrew M Childs, Richard Cleve, Enrico Deotto, Edward Farhi, Sam Gutmann, and Daniel A Spielman. Exponential algorithmic speedup by a quantum walk. In *Proceedings of the thirty-fifth annual ACM symposium on Theory of computing*, pages 59–68, 2003.
- [8] Richard Cleve, Artur Ekert, Chiara Macchiavello, and Michele Mosca. Quantum algorithms revisited. *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, 454(1969):339–354, 1998.
- [9] Don Coppersmith. An approximate Fourier Transform useful in quantum factoring. *arXiv preprint quant-ph/0201067*, 2002.
- [10] David Elieser Deutsch. Quantum computational networks. *Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences*, 425(1868):73–90, 1989.
- [11] BL Douglas and JB Wang. Efficient quantum circuit implementation of quantum walks. *Physical Review A*, 79(5):052335, 2009.
- [12] Johnnie Gray and Stefanos Kourtis. Hyper-optimized tensor network contraction. *arXiv preprint arXiv:2002.01935*, 2020.
- [13] Lov K Grover. A fast quantum mechanical algorithm for database search. In *Proceedings of the twenty-eighth annual ACM symposium on Theory of computing*, pages 212–219, 1996.
- [14] Chu Guo, Yong Liu, Min Xiong, Shichuan Xue, Xiang Fu, Anqi Huang, Xiaogang Qiang, Ping Xu, Junhua Liu, Shenggen Zheng, et al. General-purpose quantum circuit simulator with projected entangled-pair states and the quantum supremacy frontier. *Physical review letters*, 123(19):190501, 2019.
- [15] Richard A Harshman. Foundations of the PARAFAC procedure: models and conditions for an explanatory multimodal factor analysis. 1970.
- [16] Richard A Harshman. Determination and proof of minimum uniqueness conditions for PARAFAC1. *UCLA working papers in phonetics*, 22(111-117):3, 1972.
- [17] Frank L Hitchcock. The expression of a tensor or a polyadic as a sum of products. *Studies in Applied Mathematics*, 6(1-4):164–189, 1927.
- [18] Morten Kjaergaard, Mollie E Schwartz, Jochen Braumüller, Philip Krantz, Joel I-J Wang, Simon Gustavsson, and William D Oliver. Superconducting qubits: Current state of play. *Annual Review of Condensed Matter Physics*, 11:369–395, 2020.
- [19] Tamara G Kolda and Brett W Bader. Tensor decompositions and applications. *SIAM review*, 51(3):455–500, 2009.
- [20] T Loke and JB Wang. Efficient quantum circuits for Szegedy quantum walks. *Annals of Physics*, 382:64–84, 2017.
- [21] Troy D Mackay, Stephen D Bartlett, Leigh T Stephenson, and Barry C Sanders. Quantum walks in higher dimensions. *Journal of Physics A: Mathematical and General*, 35(12):2745, 2002.
- [22] Igor L Markov and Yaoyun Shi. Simulating quantum computation by contracting tensor networks. *SIAM Journal on Computing*, 38(3):963–981, 2008.
- [23] Nikolajs Nahimovs and Alexanders Rivošs. A note on the optimality of the Grover’s algorithm.

- [24] Ryosuke Okuta, Y. Unno, Daisuke Nishino, S. Hido, and Crissman. CuPy : A NumPy-compatible library for NVIDIA GPU calculations. *31st conference on neural information processing systems*, 2017.
- [25] Travis E Oliphant. A guide to NumPy. *Trelgol Publishing USA*, 2006.
- [26] Ivan V Oseledets. Tensor-train decomposition. *SIAM Journal on Scientific Computing*, 33(5):2295–2317, 2011.
- [27] Thomas E O’Brien, Brian Tarasinski, and Barbara M Terhal. Quantum phase estimation of multiple eigenvalues for small-scale (noisy) experiments. *New Journal of Physics*, 21(2):023022, 2019.
- [28] Yuchen Pang, Tianyi Hao, Annika Dugad, Yiqing Zhou, and Edgar Solomonik. Efficient 2D tensor network simulation of quantum systems. In *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis*, pages 174–187, 2020.
- [29] Giuseppe Davide Paparo and MA Martin-Delgado. Google in a quantum network. *Scientific reports*, 2:444, 2012.
- [30] Edwin Pednault, John A Gunnels, Giacomo Nannicini, Lior Horesh, Thomas Magerlein, Edgar Solomonik, Erik W Draeger, Eric T Holland, and Robert Wisnieff. Breaking the 49-qubit barrier in the simulation of quantum circuits. *arXiv preprint arXiv:1710.05867*, 2017.
- [31] Edwin Pednault, John A Gunnels, Giacomo Nannicini, Lior Horesh, and Robert Wisnieff. Leveraging secondary storage to simulate deep 54-qubit sycamore circuits. *arXiv preprint arXiv:1910.09534*, 2019.
- [32] Miklos Santha. Quantum walk based search algorithms. In *International Conference on Theory and Applications of Models of Computation*, pages 31–46. Springer, 2008.
- [33] Raqueline AM Santos. Szegedy’s quantum walk with queries. *Quantum Information Processing*, 15(11):4461–4475, 2016.
- [34] Ulrich Schollwöck. The density-matrix renormalization group in the age of matrix product states. *Annals of physics*, 326(1):96–192, 2011.
- [35] Neil Shenvi, Julia Kempe, and K Birgitta Whaley. Quantum random-walk search algorithm. *Physical Review A*, 67(5):052307, 2003.
- [36] Mario Szegedy. Quantum speed-up of Markov chain based algorithms. In *45th Annual IEEE symposium on foundations of computer science*, pages 32–41. IEEE, 2004.
- [37] Salvador Elías Venegas-Andraca. Quantum walks: a comprehensive review. *Quantum Information Processing*, 11(5):1015–1106, 2012.
- [38] Kieran Woolfe. Matrix product operator simulations of quantum algorithms. Technical report, University of Melbourne School of Physics Melbourne Australia, 2015.
- [39] Xin-Chuan Wu, Sheng Di, Emma Maitreyee Dasgupta, Franck Cappello, Hal Finkel, Yuri Alexeev, and Frederic T Chong. Full-state quantum circuit simulation by using data compression. In *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis*, pages 1–24, 2019.
- [40] Yiqing Zhou, E Miles Stoudenmire, and Xavier Waintal. What limits the simulation of quantum computers? *arXiv preprint arXiv:2002.07730*, 2020.

## A Additional Analysis for Phase Estimation

In Theorem A.1 we will show that all the intermediate states on the first register of the phase estimation circuit can be approximated by a low-rank state whose CP rank is bounded by  $\mathcal{O}(1/\epsilon)$ , and the output state fidelity is at least  $1 - \epsilon$ . We look at the CP rank of the states on the first register rather than the overall state, because the rank of the overall state is highly dependent on both  $|\psi\rangle$  and the oracle, hence is difficult to analyze.

**Theorem A.1.** *For the phase estimation circuit, if  $|\psi\rangle$  is the eigenvector of  $U$ , then all the intermediate states on the first register before the  $QFT^{-1}$  operator can be represented by a rank-1 state. In addition, all the intermediate states in the  $QFT^{-1}$  operator as well as its output state can be approximated by a low-rank state with the CP rank bounded by  $\mathcal{O}(1/\epsilon)$ , and the fidelity of all the intermediate and output states on the first register are at least  $1 - \epsilon$ .*

*Proof.* Under the assumption that  $|\psi\rangle$  is the eigenvector of  $U$ , the output state of each controlled- $U^{2^j}$  gate will have the same rank as the input. For example, the output of controlled- $U^{2^{n-1}}$  will be

$$(E_1|h\rangle) \otimes \cdots \otimes (I|\psi\rangle) + (E_2|h\rangle) \otimes \cdots \otimes (U^{2^{n-1}}|\psi\rangle) = \frac{1}{\sqrt{2}}(|0\rangle + e^{i2\pi 2^{n-1}\theta}|1\rangle) \otimes |h\rangle \otimes \cdots \otimes |\psi\rangle, \quad (\text{A.1})$$

and the first register state is

$$\frac{1}{\sqrt{2}}(|0\rangle + e^{i2\pi 2^{n-1}\theta}|1\rangle) \otimes |h\rangle \otimes \cdots \otimes |h\rangle, \quad (\text{A.2})$$

which also has rank 1. Other output states of controlled gates behave the same way, and all the intermediate states on the first register before  $QFT^{-1}$  all have rank 1.

Next we analyze the output state of  $QFT^{-1}$ . We can rewrite the expression for the state expressed in (5.5) as

$$\frac{1}{2^n} \sum_{x=0}^{2^n-1} \sum_{k=0}^{2^n-1} e^{-\frac{2\pi ik}{2^n}(x-2^n\theta)} |x\rangle = \frac{1}{2^n} \sum_{x=0}^{2^n-1} \sum_{k=0}^{2^n-1} e^{-\frac{2\pi ik}{2^n}(x-a)} e^{2\pi i\delta k} |x\rangle, \quad (\text{A.3})$$

where  $a$  is the nearest integer to  $2^n\theta$ ,  $2^n\theta = a + 2^n\delta$  and  $0 \leq |2^n\delta| \leq \frac{1}{2}$ . Let  $\alpha^{(t)}$  denote the amplitude of  $|a - t \bmod 2^n\rangle$  and  $-2^{n-1} \leq t < 2^{n-1}$ , as is shown in reference [8],  $\alpha^{(t)}$  can be expressed as

$$\alpha^{(t)} = \frac{1}{2^n} \frac{1 - e^{2\pi i 2^n (\delta + \frac{t}{2^n})}}{1 - e^{2\pi i (\delta + \frac{t}{2^n})}}, \quad (\text{A.4})$$

and the probability of outputting states that are at least  $k$ -away from  $|a\rangle$  is bounded by

$$\sum_{k \leq |t| \leq 2^{n-1}} |\alpha^{(t)}|^2 < \frac{1}{2k-1}. \quad (\text{A.5})$$

For  $\frac{1}{2k-1} \leq \epsilon$ , we need  $k = \mathcal{O}(1/\epsilon)$ . Therefore, with high fidelity  $(1-\epsilon)$ , the output state of the first register can be approximated with a state that is in the space of  $\{|x\rangle, a-k \leq x \leq a+k\}$ . The approximated state can be written as the linear combination of  $2k = \mathcal{O}(1/\epsilon)$  standard basis states, and the CP rank is bounded by  $\mathcal{O}(1/\epsilon)$ .

Let  $|\psi_a\rangle, |\psi_t\rangle$  denote the accurate output state, and the approximated low-rank output state of  $QFT^{-1}$ , respectively. Let  $|\phi_a\rangle$  denote one accurate intermediate state of  $QFT^{-1}$ . We can express  $|\psi_a\rangle = U|\phi_a\rangle$  for some unitary  $U$ . The fidelity of  $|\psi_t\rangle$  can be expressed as

$$|\langle \psi_t | \psi_a \rangle|^2 = |\langle \psi_t | U | \phi_a \rangle|^2 = |\langle \phi_a | \phi_t \rangle|^2 \geq 1 - \epsilon, \quad (\text{A.6})$$

where  $|\psi_t\rangle = U|\phi_t\rangle$ . Above equation means that all the intermediate states  $|\phi_a\rangle$  can be approximated by  $|\phi_t\rangle$ , and the fidelity will also be at least  $1 - \epsilon$ . Based on Theorem 5.1 and the discussion in Section 5.1, all the approximated intermediate states  $|\phi_t\rangle$  will have rank  $\mathcal{O}(1/\epsilon)$ . Therefore, all the intermediate states can be accurately approximated by low-rank states, and the theorem is proved.  $\square$

## B Additional Analysis for Grover's Algorithm

The rank of  $U^{(g)}|\psi\rangle$  is dependent on the implementation of  $U^{(g)}$ , which is usually regarded as a black box. For the simulation to be efficient,  $U^{(g)}|\psi\rangle$  needs to be in the CP representation, and the rank needs to be low. We will show in Theorem B.1 that the rank of the state  $U^{(g)}|\psi\rangle$  will be at most  $2(a+1)$  times the rank of  $|\psi\rangle$ , thus the rank is low consider that  $a$  is small.

**Theorem B.1.** *Consider an state  $|\psi\rangle$  in the CP representation, and the rank is  $R$ . Then there exists implementations for  $U^{(g)}$  such that it consists of  $a+1$  generalized controlled gates, and the output state  $U^{(g)}|\psi\rangle$  is also in the CP representation and the rank is at most  $2(a+1)R$ .*

*Proof.* We assume that  $U^{(d)}$  is implemented similar to what is shown in Figure 3, and  $U^{(o)}$  is implemented with  $|A| = a$  generalized controlled gates. Each generalized controlled gate flips the sign for one state  $|x\rangle$  representing an marked item. Let  $CU^{(x)}$  denote the controlled gate for  $|x\rangle = |x_1x_2\cdots x_n\rangle$ , we have

$$CU^{(x)} = I \otimes \cdots \otimes I - 2E_{x_1} \otimes \cdots \otimes E_{x_n}, \quad (\text{B.1})$$

and  $U^{(o)}$  can be expressed as

$$U^{(g)} = CU^{(y^{(1)})}CU^{(y^{(2)})}\cdots CU^{(y^{(a)})}, \quad (\text{B.2})$$

where the marked set  $A = \{y^{(1)}, \dots, y^{(a)}\}$ . (B.2) can be rewritten as

$$U^{(o)} = I \otimes \cdots \otimes I - \sum_{i=1}^a 2E_{y_1^{(i)}} \otimes \cdots \otimes E_{y_n^{(i)}}, \quad (\text{B.3})$$

which contains the summation of  $a+1$  Kronecker products. (B.3) holds since for two different standard basis  $u$  and  $v$ ,

$$(E_{u_1} \otimes \cdots \otimes E_{u_n})(E_{v_1} \otimes \cdots \otimes E_{v_n}) = O. \quad (\text{B.4})$$

Therefore, applying  $U^{(o)}$  can output a state in the CP representation with the rank being  $(a+1)R$ . Applying  $U^{(d)}$  then will at most double the rank. Therefore, the output of  $U^{(g)}$  is in the CP representation and the rank is at most  $2(a+1)R$ .  $\square$

In Theorem B.2, we will show that for any marked set that is small in size, we can approximate the intermediate states by rank-2 states, and the simulation will still output one marked state will high probability with  $\mathcal{O}(\sqrt{N})$  Grover's iterations.

**Theorem B.2.** *When  $|A| = a$  is a constant and  $a \ll N$ , one marked state will be found with high probability if all the intermediate states are approximated by a low-rank state with CP rank upper bounded by 2, with the operator  $U^{(g)}$  applied  $\mathcal{O}(\sqrt{N})$  times.*

*Proof.* Let  $|a\rangle$  denote one of the states in the marked set. We consider the case when all the intermediate approximate states are in the space  $S'$  spanned by  $|a\rangle$  and  $|h\rangle$ , whose CP ranks are at most 2. Note that the input state  $|h\rangle$  is also in the span. We will show that after applying operator  $U^{(g)}$  for  $\lfloor \frac{\pi}{4}\sqrt{N} \rfloor$  times, the output is close to  $|a\rangle$ .

Consider an input state  $|x\rangle = \alpha|a\rangle + \beta|h\rangle$ , where  $a$  is one of the target states in  $A$ . Then

$$U^{(g)}|a\rangle = (2|h\rangle\langle h| - I)(U^{(o)}|a\rangle) = |a\rangle - 2|h\rangle\langle h|a\rangle = |a\rangle - \frac{2}{\sqrt{N}}|h\rangle, \quad (\text{B.5})$$

$$U^{(g)}|h\rangle = U^{(g)}\left(\sqrt{\frac{a}{N}}|A\rangle + \sqrt{\frac{b}{N}}|B\rangle\right) = \left(1 - \frac{4a}{N}\right)|h\rangle + 2\sqrt{\frac{a}{N}}|A\rangle, \quad (\text{B.6})$$

where  $U^{(g)}|a\rangle$  is in  $S'$  while  $U^{(g)}|h\rangle$  is not. The low-rank approximation is performed on  $U^{(g)}|h\rangle$ , resulting in the (unnormalized) state  $(1 - \frac{4a}{N})|h\rangle + 2\sqrt{\frac{1}{N}}|a\rangle$ . The step of applying  $U^{(g)}|x\rangle$  and then perform the low-rank approximation can be written as a linear transformation  $G'$ , where

$$G' = \begin{bmatrix} 1 & 2\sqrt{\frac{1}{N}} \\ -2\sqrt{\frac{1}{N}} & 1 - \frac{4a}{N} \end{bmatrix}. \quad (\text{B.7})$$

The first row is applied on the  $|a\rangle$  component and the second row on  $|h\rangle$ . Let  $|b\rangle = \frac{\sqrt{N}|h\rangle - |a\rangle}{\sqrt{N-1}}$  denoting the superposition of all the states except  $|a\rangle$ . we can rewrite  $G'$  based on the basis  $|a\rangle, |b\rangle$  as follows:

$$G' = \begin{bmatrix} 1 & \frac{1}{N} \\ 0 & \frac{N-1}{N} \end{bmatrix} \begin{bmatrix} 1 & 2\sqrt{\frac{1}{N}} \\ -2\sqrt{\frac{1}{N}} & 1 - \frac{4a}{N} \end{bmatrix} \begin{bmatrix} 1 & -\frac{1}{N-1} \\ 0 & \frac{N}{N-1} \end{bmatrix}, \quad (\text{B.8})$$

and the first row is applied on the  $|a\rangle$  component and the second row on  $|b\rangle$ . Above operator can be rewritten as

$$\begin{bmatrix} 1 - \frac{2}{N} & 2\frac{\sqrt{N-1}}{N} \\ -2\frac{\sqrt{N-1}}{N} & 1 - \frac{2}{N} \end{bmatrix} + \begin{bmatrix} 0 & \frac{4-4a}{N}\sqrt{\frac{1}{N-1}} \\ 0 & \frac{4-4a}{N} \end{bmatrix}. \quad (\text{B.9})$$

Note that the first component is equal to the operator  $U^{(g)}$  for the Grover's algorithm when the marked set only has  $|a\rangle$ , and the second component comes from both  $|A| > 1$  and the low-rank approximation. When  $a \ll N$ , it can be observed that the second component can be negligible, and the approximated algorithm will output  $|a\rangle$  with high probability.  $\square$

Because the approximated operator  $G'$  in (B.8) is close to the  $U^{(g)}$  for the system with the marked set size equal to 1, the number of iterations necessary to get a marked state with high probability increases from  $\lfloor \frac{\pi}{4}\sqrt{\frac{N}{a}} \rfloor$  to  $\lfloor \frac{\pi}{4}\sqrt{N} \rfloor$ . For the system when  $a$  is unknown,  $\lfloor \frac{\pi}{4}\sqrt{N} \rfloor$  iterations need to be performed for both.