

## Low-depth quantum state preparation

Xiao-Ming Zhang,<sup>1,2</sup> Man-Hong Yung<sup>3,4,5,\*</sup> and Xiao Yuan<sup>1,†</sup>

<sup>1</sup>Center on Frontiers of Computing Studies, Department of Computer Science, Peking University, Beijing 100871, China

<sup>2</sup>Department of Physics, City University of Hong Kong, Tat Chee Avenue, Kowloon, Hong Kong SAR, China

<sup>3</sup>Department of Physics, Southern University of Science and Technology, Shenzhen 518055, China

<sup>4</sup>Shenzhen Institute for Quantum Science and Engineering, Southern University of Science and Technology, Shenzhen 518055, China

<sup>5</sup>Guangdong Provincial Key Laboratory of Quantum Science and Engineering, Southern University of Science and Technology, Shenzhen 518055, China



(Received 15 March 2021; accepted 10 November 2021; published 21 December 2021)

A crucial subroutine in quantum computing is to load the classical data of  $N$  complex numbers into the amplitude of a superposed  $n = \lceil \log_2 N \rceil$ -qubit state. It has been proven that any algorithm universally implementing this subroutine would need at least  $O(N)$  constant weight operations. However, the proof assumes that only  $n$  qubits are used, whereas the circuit depth could be reduced by extending the space and allowing ancillary qubits. Here we investigate this space-time tradeoff in quantum state preparation with classical data. We propose quantum algorithms with  $O(n^2)$  circuit depth to encode any  $N$  complex numbers using only single- and two-qubit gates, and local measurements with ancillary qubits. Different variances of the algorithm are proposed with different space and runtime. In particular, we present a scheme with  $O(N^2)$  ancillary qubits,  $O(n^2)$  circuit depth, and  $O(n^2)$  average runtime, which exponentially improves the conventional bound. While the algorithm requires more ancillary qubits, it consists of quantum circuit blocks that only simultaneously act on a constant number of qubits, and at most  $O(n)$  qubits are entangled. We also prove a fundamental lower bound  $\Omega(n)$  for the minimum circuit depth and runtime with an arbitrary number of ancillary qubits, aligning with our scheme with  $O(n^2)$ . The algorithms are expected to have wide applications in both near-term and universal quantum computing.

DOI: [10.1103/PhysRevResearch.3.043200](https://doi.org/10.1103/PhysRevResearch.3.043200)

### I. INTRODUCTION

Various quantum algorithms have been designed for solving different types of problems [1]. A critical subroutine of many quantum algorithms is to encode classical data into a superposed quantum state [2–8], which prepares a general multiqubit state with classically given amplitudes. An efficient state preparation scheme is the prerequisite of many algorithms, including quantum linear system algorithms [9,10], quantum versions of data fitting [11], principal component analysis [12], support vector machines [13], Hamiltonian simulation algorithms [14–16], quantum machine learning [17–23], etc. Theoretically, the minimal number of constant-weight operations (i.e., operations applied on no more than a constant number of qubits) to prepare an arbitrary  $N$ -dimensional or  $n = \lceil \log_2 N \rceil$ -qubit state is lower bounded by  $\Omega(N/\ln n)$  [1,24], which corresponds to the circuit depth of  $\Omega(N/(n \ln n))$ . For instance, one may construct a unitary to transform  $|0\rangle^{\otimes n}$  to the target state with only single-qubit and CNOT gates, and existing algorithms [4,5,7,8] require  $O(N)$

circuit depth, which is close to the fundamental limit. Since the complexity is linear in the dimension  $N$  or exponential in the number of qubits  $n$ , it requires a deep circuit for large  $N$ . For example, the circuit depth is already challenging for the current technology when  $N \approx 10^3$  or  $n = 10$ .

However, the proof of the lower bound considers operations on exactly  $n$  qubits, and one may trade the circuit depth (time) with ancillary qubits (space). Along these lines, quantum circuits with  $O(n^2)$  depth have been proposed to encode binary vectors [25] and general nonbinary vectors [26,27] into special types of entangled states. The key idea is to apply operations on  $N$  qubits in parallel so that the circuit depth is polylogarithmic in  $N$ . In Ref. [26], the method has also been applied for improving quantum machine-learning algorithms. Nevertheless, the output quantum state is encoded with  $N$  qubits, which is exponentially larger than  $n$  and is in a complicated entangled basis of all the  $N$  qubits, which may not be universally usable as the input to other quantum algorithms. While there are other methods with logarithmic costs, including controlled-rotation-based [3,28] and Grover-oracle-based [29] methods, they require global unitaries or global oracles acting on all qubits, which is challenging based on current technologies. Therefore, it remains an open question whether it is possible to more efficiently and directly prepare a general  $N$ -dimensional ( $n$ -qubit) quantum state with constant-weight operations and a shallow circuit depth.

In this work, we address this problem by introducing probabilistic quantum state preparation algorithms. We consider the task of preparing an  $N$ -dimensional ( $n$ -qubit) state, and

\*yung@sustech.edu.cn

†xiaoyuan@pku.edu.cn

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/) license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

we introduce several quantum algorithms that use circuits with polylogarithmic depth  $O(n^2)$ . With different numbers of ancillary qubits, the algorithms have different success probabilities, which could be enhanced to  $O(1)$  with a runtime that is inversely proportional to the success probability. As a result, the sequential algorithm uses  $O(n)$  ancillary qubits with an average runtime  $O(N^2)$ , and the parallel algorithms uses more ancillary qubits with a smaller average runtime. Specifically, the extreme parallel algorithm has an average runtime of  $O(n^2)$  with  $O(N^2)$  ancillary qubits. Note that for all the proposed algorithms, one only needs to maintain entanglement of at most  $O(n)$  qubits. Our results thus show the space-time tradeoff in quantum state preparation. Moreover, we have shown that fundamentally, the circuit depth and runtime is lower bounded by  $\Omega(n)$  even with an arbitrarily large amount of ancillary qubits, which is comparable to our result of  $O(n^2)$ .

### II. FRAMEWORK

We first introduce the task of quantum state preparation. Given a vector  $\mathbf{u} := [u_0, u_1, \dots, u_{N-1}] \in \mathbb{C}^N$  of  $N$  complex numbers satisfying  $\|\mathbf{u}\|_2 = 1$ , we consider the preparation of the  $n$ -qubit state

$$|\psi(\mathbf{u})\rangle := \sum_{i=0}^{N-1} u_i |n, i\rangle, \quad (1)$$

where  $|n, i\rangle$  is the  $n$ -qubit binary representation of  $i$ . For example,  $|3, 7\rangle = |111\rangle$ ,  $|3, 6\rangle = |110\rangle$ , and  $|4, 7\rangle = |0111\rangle$ . Here, the state  $|\psi(\mathbf{u})\rangle$  is also called the *amplitude encoding* [30–35] of the vector  $\mathbf{u}$ , and it serves as our target state. The reason for including an additional index  $n$  is that there are states with different numbers of qubits involved in our algorithms. The main problem we consider is as follows:

Given an arbitrary quantum state described in Eq. (1), find a preparations method with constant-weight operations and a polylogarithmic circuit depth.

In addition, we will also discuss the tradeoff between circuit-depth, number of ancillary qubits, and preparation fidelity.

To prepare  $|\psi(\mathbf{u})\rangle$ , we consider a resized vector  $\mathbf{v} := \mathbf{u}/\max(|u_i|)$  and define the *label encoding state* of  $\mathbf{v}$  with  $n + 1$  qubits (below, quantum states may be represented up to a normalization factor),

$$|\mathbf{v}\rangle := \sum_{i=0}^{N-1} |n, i\rangle |v_i\rangle = \sum_{i=0}^{N-1} |n, i\rangle [v_i|0\rangle + (1 - v_i)|1\rangle], \quad (2)$$

where  $|n, i\rangle$  and  $|v_i\rangle = v_i|0\rangle + (1 - v_i)|1\rangle$  represent the  $n$ -qubit label and the value single qubit, respectively. Note that if we project the value qubit to  $|0\rangle$  and trace it out, we can probabilistically obtain the target state  $|\psi(\mathbf{u})\rangle$ . Thus we focus first on the preparation of the label encoding state.

### III. POSITIVE LABEL STATE PREPARATION

We first consider the special case with positive amplitudes  $\mathbf{v} \in [0, 1]^N$ . Our algorithm is based on the following result about concatenating two label encoding states.

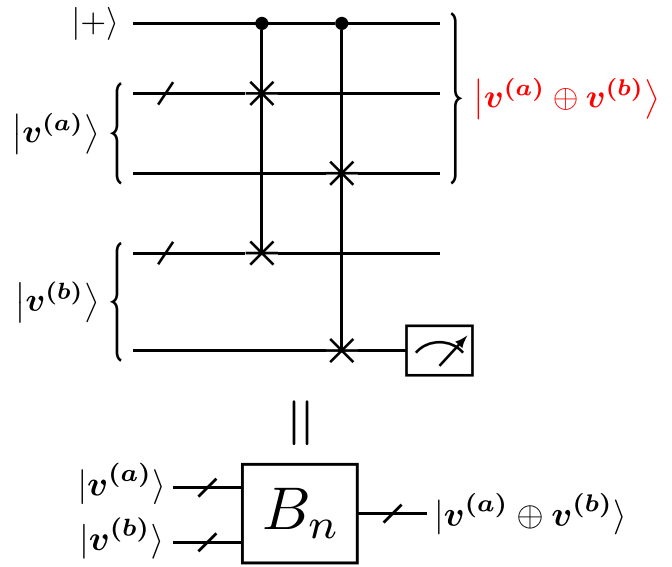


FIG. 1. Low-depth concatenation circuit for preparing the label encoding state  $|\mathbf{v}^{(a)} \oplus \mathbf{v}^{(b)}\rangle$  with input  $|\mathbf{v}^{(a)}\rangle \otimes |\mathbf{v}^{(b)}\rangle$  and ancillary qubit  $|+\rangle$ . Meter represents projecting qubits to state  $|+\rangle$ , and red represents the target output state. See Appendix A for details.

*Result 1.* Given two quantum states  $|\mathbf{v}^{(a)}\rangle$  and  $|\mathbf{v}^{(b)}\rangle$  with  $\mathbf{v}^{(a)}, \mathbf{v}^{(b)} \in [0, 1]^N$ , there exists an  $O(n)$  depth concatenation circuit, such that the state  $|\mathbf{v}^{(a)} \oplus \mathbf{v}^{(b)}\rangle$  can be obtained with a probability larger than  $1/2$ .

We have defined  $\mathbf{v}^{(a)} \oplus \mathbf{v}^{(b)} = [v_0^{(a)}, \dots, v_{N-1}^{(a)}, v_0^{(b)}, \dots, v_{N-1}^{(b)}]$ . The *concatenation circuit* is shown in Fig. 1, where we perform a joint controlled-swap operation on each pair of qubits for  $|\mathbf{v}^{(a)}\rangle$  and  $|\mathbf{v}^{(b)}\rangle$  with a control ancillary qubit initialized in  $|+\rangle = 1/\sqrt{2}(|0\rangle + |1\rangle)$ . The state then becomes

$$\frac{1}{\sqrt{2}}(|0\rangle|\mathbf{v}^{(a)}\rangle \otimes |\mathbf{v}^{(b)}\rangle + |1\rangle|\mathbf{v}^{(b)}\rangle \otimes |\mathbf{v}^{(a)}\rangle), \quad (3)$$

where  $\otimes$  is the Kronecker product. Next, the key step is to disentangle the last  $(n + 1)$  qubits by projecting the last value qubit to  $|+\rangle$ . The success probability of the projection satisfies  $p_+ \geq 1/2$  (see Appendix A). After the projection, the last  $n + 1$  qubits are disentangled with the remaining  $n + 2$  qubits, and the full quantum state is given by

$$(|0\rangle|\mathbf{v}^{(a)}\rangle + |1\rangle|\mathbf{v}^{(b)}\rangle) \otimes |\mathbf{v}^{\text{uni}}\rangle = |\mathbf{v}^{(a)} \oplus \mathbf{v}^{(b)}\rangle \otimes |\mathbf{v}^{\text{uni}}\rangle,$$

where we have defined  $\mathbf{v}^{\text{uni}} = [1/2, 1/2, \dots, 1/2]$ . Note that  $|\mathbf{v}^{(a)} \oplus \mathbf{v}^{(b)}\rangle = \sum_{i=0}^{2N-1} |n + 1, i\rangle [v_i|0\rangle + (1 - v_i)|1\rangle]$ , where  $v_i = v_i^{(a)}$  for  $i < N$  and  $v_i = v_{i-N}^{(b)}$  for  $i \geq N$ . By tracing out  $|\mathbf{v}^{\text{uni}}\rangle$ , the label encoding state of the concatenated vectors  $|\mathbf{v}^{(a)} \oplus \mathbf{v}^{(b)}\rangle$  is obtained. Because  $|\mathbf{v}^{(a)}\rangle$  and  $|\mathbf{v}^{(b)}\rangle$  are  $(n + 1)$  qubit states, and each control swap gate can be realized with a constant number of single- and two-qubit gates, the concatenation circuit has  $O(n)$  circuit depth.

The sequential scheme works by applying the concatenation circuit with a divide-and-conquer strategy (see also Algorithm 1). As shown in Fig. 2(a), we sequentially prepare the  $(i + 1)$ -qubit label encoding state via the concatenation circuit. All ancillary qubits are disentangled after the measurement and can be reused, so only  $O(i)$  ancillary qubits

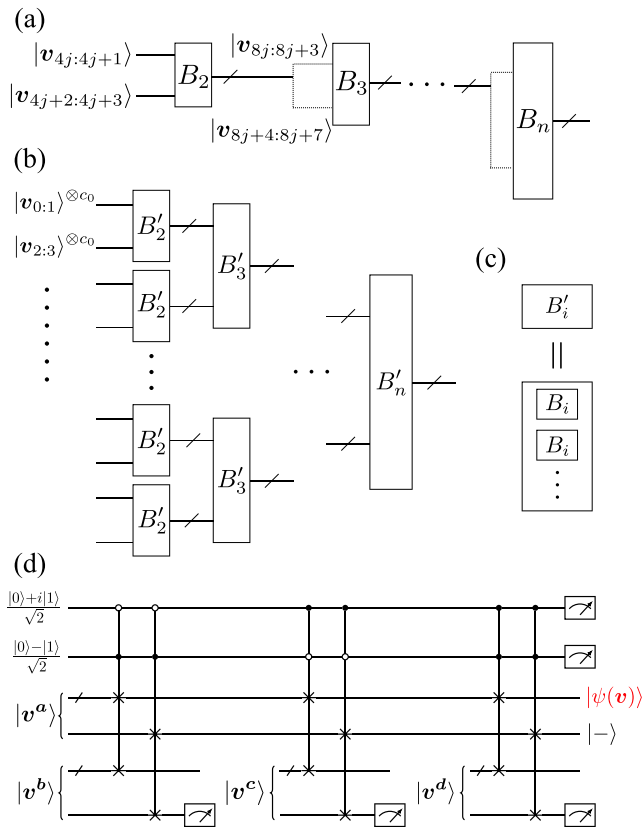


FIG. 2. (a),(b) Low-depth quantum circuit for preparing a positive label encoding state for (a) sequential parallel preparation and (b) parallel preparation. (c) A parallel concatenation circuit  $B'_i$  contains multiple (nonparallel) concatenation circuits  $B_i$ . (d) Circuit for preparing arbitrary states  $|\psi(\bar{v})\rangle$  with four positive label encoding states. Red represents the target output state.

are required at each step. The sequential scheme thus requires  $O(i)$  ancillary qubits and  $O(n)$  circuit depth. Meanwhile, we need to repeat the concatenation circuit several times to obtain the output state with constant probabilities. We denote the average runtime for preparing  $(i + 1)$  qubit positive label encoding states as  $T_{\text{pos}}(i)$ . We also assume that each layer of quantum gates takes a constant operation time. Because the success probability of each concatenation circuit is larger than  $1/2$ , we have  $T_{\text{pos}}(i) \leq 2[2T_{\text{pos}}(i - 1) + k_1 i + k_0]$ . Here,  $k_1$  characterizes the runtime for a single control swap gate, and  $k_0$  characterizes the runtime for processes that are independent of  $i$ , such as detection time and latent time. We show that  $T_{\text{pos}}(n) \leq O(N^2)$ .

The runtime can be improved by parallelization. First, two input states of the concatenation circuit (see Fig. 1) can be prepared in parallel. Second, one can prepare sufficient copies of the input state and then perform the concatenation circuit for multiple pairs of input states in parallel. In this way, the success probability of the projection (with at least one successful transformation) will be much higher, and the total runtime could be reduced dramatically. In practical implementation, this can help to reduce the decoherent error significantly. In Fig. 2(c), we show the *parallel concatenation circuit*  $B'_i$  for preparing  $(i + 1)$ -qubit label encoding

**Algorithm 1:**  $f_{\text{seq}}(\mathbf{x})$

- 
- 1: **If**  $\mathbf{x}$  is two-dimensional:
  - 2:   prepare  $|\mathbf{x}\rangle$  with the unitary preparation method
  - 3:   **Output**  $|\mathbf{x}\rangle$
  - 4: **Else:**
  - 5:   let  $\mathbf{x}^{(a)} \oplus \mathbf{x}^{(b)} = \mathbf{x}$
  - 6:   prepare  $|\mathbf{x}^{(a)}\rangle$  with  $f_{\text{seq}}(\mathbf{x}^{(a)})$
  - 7:   prepare  $|\mathbf{x}^{(b)}\rangle$  with  $f_{\text{seq}}(\mathbf{x}^{(b)})$
  - 8:   perform transformation  $|\mathbf{x}^{(a)}\rangle \otimes |\mathbf{x}^{(b)}\rangle \rightarrow |\mathbf{x}\rangle$  (with success probability on less than  $1/2$ )
  - 9:   **If** the transformation in line 8 fails:
  - 10:     go to line 6
  - 11:   **Else:**
  - 12:     **Output**  $|\mathbf{x}\rangle$
- 

states, which receives state  $|\mathbf{v}^{(a)}\rangle^{\otimes c_a} \otimes |\mathbf{v}^{(b)}\rangle^{\otimes c_b}$ . One performs the concatenation circuit  $c_{\text{min}} = \min\{c_a, c_b\}$  times in parallel, with a total of  $c'$  successful trials, and we obtain the output state  $|\mathbf{v}^{(a)} \oplus \mathbf{v}^{(b)}\rangle^{\otimes c'}$ . Note that  $c'$  follows a binomial distribution  $c' \sim B(c_{\text{min}}, p_+)$  with  $p_+ \geq 1/2$ .

The parallel preparation of an  $(n + 1)$ -qubit label encoding state is shown in Fig. 2(b), where we have denoted  $\mathbf{v}_{i:j} = [v_i, v_{i+1}, \dots, v_j]$ . We prepare  $c_0$  copies of the low dimensional label encoding states, i.e.,  $|\mathbf{v}_{0:1}\rangle^{\otimes c_0}, |\mathbf{v}_{2:3}\rangle^{\otimes c_0}, \dots$ , which are concatenated recursively to state  $|\mathbf{v}\rangle$ . Whenever the parallel concatenation circuit has zero copies of output, we repeat the preparation of the input state of the corresponding block (see also Algorithm 2). The average runtime and space complexity depend on  $c_0$ . For example, when  $c_0 = \lceil N + N^{3/4} \rceil$ , the ancillary qubit number scales as  $O(N^2)$ , while  $T_{\text{pos}}(n)$  scales polylogarithmically as  $O(n^2)$  (see Appendix B). When  $c_0 = 1$ , we need  $O(N)$  ancillary qubits, and we numerically find that  $T_{\text{pos}}(n) = O(N^{1.52})$  in the worst case, i.e.,  $p_+ = 0.5$  (see Appendix B). We note that although more ancillary qubits are needed in the parallel schemes, only the entanglement among at most  $(n + 1)$  qubits is required, and all ancillary qubits can be reused after the preparation.

With the label encoding state, the amplitude encoding state  $|\psi(\mathbf{v})\rangle$  can be obtained by projecting the value qubit to  $|0\rangle$

**Algorithm 2:**  $f_{\text{para}}(\mathbf{x}, c_0)$

- 
- 1: **If**  $\mathbf{x}$  is two-dimensional:
  - 2:   prepare  $c_0$  copies of  $|\mathbf{x}\rangle$  with the unitary preparation method in parallel
  - 3:   **Output**  $|\mathbf{x}\rangle^{\otimes c_0}$
  - 4: **Else:**
  - 5:   let  $\mathbf{x}^{(a)} \oplus \mathbf{x}^{(b)} = \mathbf{x}$
  - 6:   query  $f_{\text{para}}(\mathbf{x}^{(a)}, c_0)$  and  $f_{\text{para}}(\mathbf{x}^{(b)}, c_0)$  in parallel, get return  $|\mathbf{x}^{(a)}\rangle^{\otimes c_a}$  and  $|\mathbf{x}^{(b)}\rangle^{\otimes c_b}$
  - 7:   define  $c_{\text{min}} = \min\{c_a, c_b\}$
  - 8:   perform transformation  $|\mathbf{x}^{(a)}\rangle \otimes |\mathbf{x}^{(b)}\rangle \rightarrow |\mathbf{x}\rangle$  for  $c_{\text{min}}$  times *in parallel*, with  $c$  trials success
  - 9:   **If**  $c = 0$ :
  - 10:     go to line 6
  - 11:   **Else:**
  - 12:     **Output**  $|\mathbf{x}\rangle^{\otimes c}$
-

with probability  $p_s$ , the value of which will be discussed in the Sec. IV A. In this way, an arbitrary quantum state [Eq. (1)] with real amplitudes  $u_i$  can be prepared with a low circuit depth. We summarize the result as follows:

*Result 2.* With probability  $p_s$ , an arbitrary  $n$ -qubit quantum state with real amplitudes can be prepared via the sequential and parallel algorithms with  $O(n^2)$  depth of single-qubit gates, two-qubit gates, and local measurements.

We note that the runtime  $T$  of preparing the state with constant probabilities is  $O(T_{\text{pos}}/p_s)$ , and we will shortly discuss how to bound it in the general case.

The sequential and parallel algorithms also work for the general case with complex amplitudes. However, the success probability of the concatenation circuit with two arbitrary vectors can only be lower bounded by  $p_+ \geq 1/10$ , instead of  $p_+ \geq 1/2$  for the positive amplitudes case. Therefore, the runtime or the number of qubits will be significantly increased (Appendix A).

The pseudocode of sequential and parallel preparation methods for preparing positive quantum label encoding states are provided as Algorithms 1 and 2, where  $\mathbf{x}$  represents the (unnormalized) amplitude of the target state, and  $f_{\text{seq}}(\mathbf{x})$ ,  $f_{\text{para}}(\mathbf{x}, c_0)$  are the target output quantum states. The quantum devices mainly realize lines 2 and 8 in both algorithms.

#### IV. ARBITRARY STATE PREPARATION

Now we propose an alternative strategy to prepare the label encoding state with arbitrary complex amplitudes. We rewrite  $\mathbf{v}$  as the combination of four positive vectors,  $\mathbf{v} = \mathbf{v}^a - \mathbf{v}^b + i\mathbf{v}^c - i\mathbf{v}^d$ , whose elements are defined as  $v_i^a = \max(\text{Re}(v_i), 0)$ ,  $v_i^b = \max(-\text{Re}(v_i), 0)$ ,  $v_i^c = \max(\text{Im}(v_i), 0)$ , and  $v_i^d = \max(-\text{Im}(v_i), 0)$ , respectively. First, the four positive label encoding states  $|\mathbf{v}^a\rangle$ ,  $|\mathbf{v}^b\rangle$ ,  $|\mathbf{v}^c\rangle$ , and  $|\mathbf{v}^d\rangle$  could be prepared with the above scheme. Then, we introduce two ancillary qubits prepared in states  $(|0\rangle + i|1\rangle)/\sqrt{2}$  and  $(|0\rangle - |1\rangle)/\sqrt{2}$ . The entire system is described by

$$(|00\rangle - |01\rangle + i|10\rangle - i|11\rangle) \otimes |\mathbf{v}^{abcd}\rangle, \quad (4)$$

where we have used the abbreviation  $|\mathbf{v}^{abcd}\rangle \equiv |\mathbf{v}^a\rangle \otimes |\mathbf{v}^b\rangle \otimes |\mathbf{v}^c\rangle \otimes |\mathbf{v}^d\rangle$ . To obtain  $|\mathbf{v}\rangle$ , we perform three sets of controlled-controlled-swap gates, which swap  $|\mathbf{v}^a\rangle$  and one of the states among  $|\mathbf{v}^b\rangle$ ,  $|\mathbf{v}^c\rangle$ , and  $|\mathbf{v}^d\rangle$ , with two ancillary qubits as control qubits. The corresponding quantum circuit is shown in Fig. 2(d), where the hollow nodes represent control on  $|0\rangle$ , and the solid nodes represent control on  $|1\rangle$ . The state then becomes

$$|00\rangle|\mathbf{v}^{abcd}\rangle - |01\rangle|\mathbf{v}^{bacd}\rangle + i|10\rangle|\mathbf{v}^{cbad}\rangle - i|11\rangle|\mathbf{v}^{dabc}\rangle. \quad (5)$$

The above operations require a total of  $3(n+1)$  control-control-swap gates, each of which can be decomposed to constant numbers of single- and two-qubit gates, so the corresponding circuit depth is  $O(n)$ .

In the next step, we project two ancillary qubits and the label qubits of the last three label encoding states to  $|+\rangle$ . If the projection succeeds (the success probability  $p'_s$  will be discussed later), the quantum state becomes

$$|+\rangle^{\otimes 2}(|\mathbf{v}^a\rangle - |\mathbf{v}^b\rangle + i|\mathbf{v}^c\rangle - i|\mathbf{v}^d\rangle) \otimes |\mathbf{v}^{\text{uni}}\rangle^{\otimes 3}. \quad (6)$$

Since  $|\mathbf{v}^a\rangle - |\mathbf{v}^b\rangle + i|\mathbf{v}^c\rangle - i|\mathbf{v}^d\rangle = \sqrt{2}|\psi(\mathbf{v})\rangle \otimes |-\rangle$ , we can trace out  $|+\rangle^{\otimes 2}$  and  $|-\rangle \otimes |\mathbf{v}^{\text{uni}}\rangle^{\otimes 3}$  to have the target state  $|\psi(\mathbf{v})\rangle$ . Together with Result 2, we have the following result:

*Result 3.* With probability  $p'_s$ , an arbitrary  $n$ -qubit quantum state can be prepared via the sequential and parallel algorithms with  $O(n^2)$  depth of single-qubit gates, two-qubit gates, and local measurements.

We note that the average runtime is proportional to  $T_{\text{pos}}$  for preparing each  $|\mathbf{v}^a\rangle$ ,  $|\mathbf{v}^b\rangle$ ,  $|\mathbf{v}^c\rangle$ , or  $|\mathbf{v}^d\rangle$ , divided by the projection success probability  $p'_s$ , i.e.,  $O(T_{\text{pos}}/p'_s)$ . Next we show how to estimate the success probabilities.

#### A. Projection success probability

To exactly prepare the amplitude encoding state  $|\psi(\mathbf{u})\rangle$ , the projection probabilities  $p_s$  (for positive data) and  $p'_s$  (for complex data) are both lower bounded by  $\Omega(\sum_i |u_i|^2 / \max(|u_i|^2)N)$ . The worst-case lower bound is  $\Omega(1/N)$  and it could be tightened with a detailed analysis. Denoting  $u_i = |a_i|/\sqrt{\sum_i |a_i|^2}$  for positive data or  $u_i = a_i/\sqrt{\sum_i |a_i|^2}$  for general complex data, we consider that the classical data  $\mathbf{u}$  are randomly generated in two ways.

(i) For the first way, we let  $a_i = b_i e^{i\phi_i}$ , and we uniformly generate each  $b_i$  from  $[-1, 1]$ , and  $\phi_i$  from  $[0, \pi]$ .

(ii) For the second way, we let  $a_i = a_i^{(r)} + ia_i^{(m)}$ , and we generate each  $a_i^{(r)}$ ,  $a_i^{(m)}$  according to the standard normal distribution  $\mathcal{N}(0, 1)$ .

The first way corresponds to the case where the classical data, i.e., each  $u_i$ , are uniformly random; the second way corresponds to the case where the state vector  $|\psi(\mathbf{u})\rangle$  is uniformly random in the Hilbert space [36]. Then we can lower bound the projection probability from the Chernoff bound as follows (see Appendix C 1).

*Result 4.* With failure probability  $\delta \in (0, 1)$ , the projection probabilities are lower bounded by

$$\begin{aligned} \text{case 1: } & p_s, p'_s \geq \Omega(\delta^{1/N}), \\ \text{case 2: } & p_s, p'_s \geq \Omega(\delta^{1/N} / \ln(N/\delta)). \end{aligned} \quad (7)$$

Therefore, for fixed  $\delta$ , the projection probabilities could be lower bounded by a constant for case 1 and by  $\Omega(1/n)$  for case 2 given sufficiently large  $N$ .

We can further improve the projection probabilities for case 2 by allowing approximate state preparation. We introduce a cutoff value  $u_{\text{cut}}$  and define  $\tilde{v}_i \equiv \arg(u_i) \min(|u_i|/u_{\text{cut}}, 1)$ . After preparing  $|\psi(\tilde{\mathbf{v}})\rangle$  with  $\tilde{\mathbf{v}} = [\tilde{v}_0, \tilde{v}_1, \dots, \tilde{v}_{N-1}]$ , we can achieve the preparation fidelity  $F \equiv |\langle \psi(\tilde{\mathbf{v}}) | \psi(\mathbf{u}) \rangle|^2 \geq 1 - \varepsilon_{\text{th}}$  by appropriately choosing the cutoff  $u_{\text{cut}}$ . Note that a perfect preparation  $F = 1$  corresponds to  $u_{\text{cut}} = \max(|u_i|)$ . In the worst cases, we have  $p_s \geq \text{mean}(|\tilde{v}_i|^2)$  and  $p'_s \geq \text{mean}(|\tilde{v}_i|^2)/64$  (see Appendix C). These values decrease with  $u_{\text{cut}}$ , whereas the preparation fidelity  $F$  increases with  $u_{\text{cut}}$ , indicating a tradeoff between projection probability and fidelity (Appendix C 2). By setting  $u_{\text{cut}}$  appropriately,  $p_s$  ( $p'_s$ ) has a logarithmic relation with  $\varepsilon_{\text{th}}$  and  $\delta$ . We summarize our results as follows:

*Result 5.* With threshold infidelity  $\varepsilon_{\text{th}} \in (0, 1)$  and failure probability  $\delta \in (0, 1)$ , the projection probabilities of obtaining the final quantum states with fidelities  $F \geq 1 - \varepsilon_{\text{th}}$  are lower

TABLE I. Comparison of different state preparation methods. Depth, circuit depth; Runtime, circuit runtime  $\times$  repetitions; Qubits, total number of qubits; Parallel-1 and -2 correspond to parallel preparation with  $c_0 = \lceil N + N^{3/4} \rceil$  and  $c_0 = 1$ . The average runtime for the parallel-2 method is estimated with numerical simulation.

	Depth	Runtime	Qubits
Unitary [4,5]	$O(N)$	$O(N)$	$O(n)$
Sequential	$O(n^2)$	$O(N^2)$	$O(n)$
Parallel-1	$O(n^2)$	$O(n^2)$	$O(N^2)$
Parallel-2	$O(n^2)$	$O(N^{1.52})$	$O(N)$

bounded by

$$p_s, p'_s \geq \Omega\left(\frac{\delta^{2/N}}{\ln \delta^{-1} \ln(\delta^{-1} \varepsilon_{\text{th}}^{-1})}\right) \quad (8)$$

if the quantum states are sampled according to case 2.

Recall that the total runtime  $T$  scales as  $O(T_{\text{pos}}/p_s)$  or  $O(T_{\text{pos}}/p'_s)$ . Combining the above results, we summarize the circuit depth, runtime, and number of qubits in Table I.

### B. Space-time tradeoff

We note that all three listed algorithms have a circuit depth  $O(n^2)$  with a different runtime and number of qubits. We can see that the runtime decreases with more qubits, indicating a space-time tradeoff in quantum state preparation. Furthermore, we change  $c_0$  for parallel preparation such that there are a total of  $\lceil N^{\beta_q} \rceil$  qubits with  $1 \leq \beta_q < 2$ , and we assume the average runtime is in the form of  $T = O(N^{\beta_t})$ . We numerically estimate the exponents  $\beta_t$  for different  $\beta_q$ . As shown in Fig. 3, the exponents  $\beta_t$  decrease rapidly with larger  $\beta_q$ . Note that with  $\beta_q$ ,  $T$  does not follow the polynomial scaling, which is consistent with our analytical estimation.

Moreover, it can be noticed that both Algorithms 1 and 2 begin with the direct preparation of label encoding states of two-dimensional vectors. One may begin with the direct preparation (e.g., with methods in Ref. [5]) of higher-dimensional states instead. In this way, the qubit numbers can be reduced, at the cost of increasing the runtime. In Appendix D, we further discuss this tradeoff with different starting state vector dimensions.

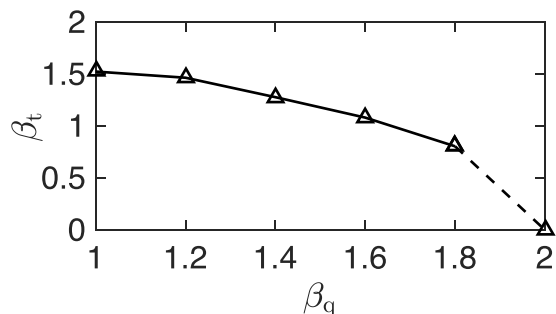


FIG. 3. Space-time tradeoff in state preparation. The relation between time  $\lceil N^{\beta_t} \rceil$  with exponent  $\beta_t$ , and space  $O(N^{\beta_q})$  with exponent  $\beta_q$  for different parallel preparation schemes. When  $\beta_q = 2$ ,  $T$  no longer follows polynomial scaling.

### C. Circuit depth and runtime lower bound

Here, we introduce a fundamental lower bound of both the circuit depth and the runtime for quantum state preparation (Appendix E). If only constant-weight operations are allowed and  $\mathbf{u}$  is stored classically, we have the following result.

*Result 6.* Preparing an arbitrary quantum state from classical amplitudes  $\mathbf{u}$  requires at least a circuit depth and runtime  $\Omega(n)$ .

The basic idea is that when  $\mathbf{u}$  is stored classically, one requires at least  $2^n$  bits to store it. To “compress” the information spread over  $2^n$  bits, at least  $\Omega(n)$  layers of the local operations circuit are required. This lower bound is independent of the number of ancillary qubits, and measurement and postselections are also taken into consideration. We note that a similar result has also been derived in Ref. [37]. Our work gives an explicit construction with circuit depth and runtime  $O(n^2)$ , which is comparable to the lower bound.

## V. DISCUSSION

We have demonstrated several protocols to prepare an arbitrary  $N$ -dimensional quantum state with  $O([\log_2(N)]^2)$  circuit depth, a different number of ancillary qubits, and a different runtime. A comparison of our methods to existing methods has been summarized in Table I. We also discuss the space-time tradeoff of the parallel preparation. In addition to the low-depth nature, there are other advantages of our work. First, our methods only require to simultaneously maintain entangled states of at most  $O(n)$  qubits, and the rest of the ancillary qubits are prepared in a separate state. Second, our methods have a much weaker requirement on the circuit programmability, since most parts of the circuit are fixed except for the first few layers. Moreover, our methods do not require heavy classical computation to compile the circuit, which typically takes a time of  $O(N)$  for unitary state preparation. We note that very recently, another important protocol with a logarithmic circuit depth has been proposed [37]. While the method in Ref. [37] requires fewer ancillary qubits, the circuit compiling procedure is more time-consuming and more complicated.

Our method can be generalized to the decomposition with discrete quantum gate set cases. We take the fault-tolerant elementary gate set {CNOT,  $H$ ,  $T$ } [38] as an example. The controlled-swap gate can be implemented with CNOT gate and single-qubit gates. According to Ref. [38], an arbitrary single-qubit gate can be approximated to accuracy  $\varepsilon$  using  $O(\ln 1/\varepsilon)$  number of  $H$  and  $T$  gates, so  $O(\ln 4^n/\varepsilon)$  gates lead to an accuracy  $O(\varepsilon/4^n)$ . Therefore, with a total circuit depth  $O(n^2 \ln(4^n/\varepsilon))$ , an  $n$ -qubit quantum state can be approximated to accuracy  $\varepsilon$ .

There are several open questions to be addressed. First, the preparation time could be much longer in the worse case. For example, if  $\mathbf{v}$  is a sparse vector with only a constant number of nonzero elements and bounded values, the success rate  $p'_s$  of the projection in Eqs. (5) and (6) decreases linearly with  $N$  and the total runtime will be  $N$  times larger. For a too small success rate, interesting future work will be to design alternative state preparation methods that exploit the sparsity and the structure of the amplitudes. Second, it is currently unclear if our

methods are optimal in terms of either circuit depth or runtime. While a lower bound has been derived in Result 6, closing the gap between  $\Omega(n)$  and  $O(n^2)$  would be compelling for both theoretical and practical purposes. Moreover, the noise robustness of our method under different noise models is also an interesting question. Finally, it is interesting to investigate applications of our methods in existing quantum algorithms and to study their performance with noisy intermediate-scaled quantum hardware [39–44]. The state preparation is a crucial subroutine for many near-term quantum algorithms [45–50], and the quantum advantage [41,51–54] could be expected with a robust, efficient, and general quantum state preparation protocol.

**ACKNOWLEDGMENTS**

We thank Tianyang Tao for helpful discussions. This work is supported by the Open Project of Shenzhen Institute of Quantum Science and Engineering (Grant No. SIQSE202008), Natural Science Foundation of Guangdong Province (Grant No. 2017B030308003), the Key R&D Program of Guangdong province (Grant No. 2018B030326001), the Science, Technology and Innovation Commission of Shenzhen Municipality (Grants No. JCYJ20170412152620376, No. JCYJ20170817105046702, and No. KYTDPT20181011104202253), National Natural Science Foundation of China (Grants No. 11875160 and No. U1801661), the Economy, Trade and Information Commission of Shenzhen Municipality (Grant No. 201901161512), and Guangdong Provincial Key Laboratory (Grant No. 2019B121203002).

**APPENDIX A: POSITIVE LABEL ENCODING STATES**

Here, we provide more details on the concatenation circuit. We will provide the proof of Result 1, and discuss the success probability for the general complex vectors.

Initially, we are given the product of an extra qubit at state  $|+\rangle = (|0\rangle + |1\rangle)/\sqrt{2}$  and the label encoding states of  $\mathbf{v}^{(a)}$  and  $\mathbf{v}^{(b)}$ . The initial state is given by (up to a normalization factor)

$$|+\rangle|\mathbf{v}^{(a)}\rangle|\mathbf{v}^{(b)}\rangle. \tag{A1}$$

According to the definition,  $\mathbf{v}^{(a)}, \mathbf{v}^{(b)}$  satisfies  $v_i^{(a)}, v_i^{(b)} \in \mathbb{C}$  and  $|v_i^{(a)}|, |v_i^{(b)}| \leq 1$ . Our goal is to transform Eq. (A1) to  $|\mathbf{v}^{(a)} \oplus \mathbf{v}^{(b)}\rangle = |0\rangle|\mathbf{v}^{(a)}\rangle + |1\rangle|\mathbf{v}^{(b)}\rangle$ . As described in the main text, we first apply a set of controlled-swap gates on  $|\mathbf{v}^{(a)}\rangle$  and  $|\mathbf{v}^{(b)}\rangle$  with the extra qubit as the control gate (see Fig. 4). The (unnormalized) quantum state then becomes

$$\begin{aligned} |\Psi_0\rangle &= \frac{1}{\sqrt{2}}|0\rangle|\mathbf{v}^{(a)}\rangle \oplus |\mathbf{v}^{(b)}\rangle + \frac{1}{\sqrt{2}}|1\rangle|\mathbf{v}^{(b)}\rangle \oplus |\mathbf{v}^{(a)}\rangle \\ &= \frac{1}{\sqrt{2}}|0\rangle|\mathbf{v}^{(a)}\rangle \sum_{i=0}^{N-1} |n, i\rangle|v_i^{(b)}\rangle \\ &\quad + \frac{1}{\sqrt{2}}|1\rangle|\mathbf{v}^{(b)}\rangle \sum_{i=0}^{N-1} |n, i\rangle|v_i^{(a)}\rangle. \end{aligned} \tag{A2}$$

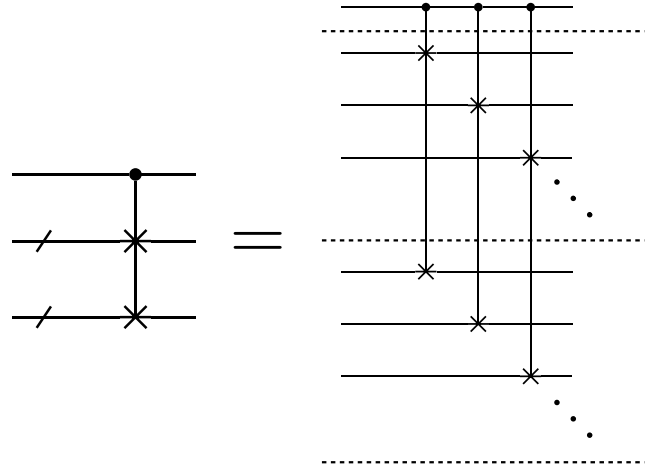


FIG. 4. Abbreviation of the control swap gate set.

We then project the last qubit (at state  $|v_i^{(a)}\rangle$  or  $|v_i^{(b)}\rangle$ ) to state  $|+\rangle$ . This is equal to applying the projection operator  $M = \mathbb{I}_{2^{2n+2}} \otimes |+\rangle\langle +|$  to Eq. (A2). Because  $\langle +|v_i^{(a)}\rangle = \langle +|v_i^{(b)}\rangle = 1/\sqrt{2}$ , we have

$$\begin{aligned} |\Psi_1\rangle &= M|\Psi_0\rangle \\ &= M \left( \frac{1}{\sqrt{2}}|0\rangle|\mathbf{v}^{(a)}\rangle \sum_{i=0}^{N-1} |n, i\rangle|v_i^{(b)}\rangle \right. \\ &\quad \left. + \frac{1}{\sqrt{2}}|1\rangle|\mathbf{v}^{(b)}\rangle \sum_{i=0}^{N-1} |n, i\rangle|v_i^{(a)}\rangle \right) \\ &= \frac{1}{2}|0\rangle|\mathbf{v}^{(a)}\rangle \sum_{i=0}^{N-1} |n, i\rangle|+\rangle + \frac{1}{2}|1\rangle|\mathbf{v}^{(b)}\rangle \sum_{i=0}^{N-1} |n, i\rangle|+\rangle \\ &= \frac{\sqrt{2}}{2}(|0\rangle|\mathbf{v}^{(a)}\rangle + |1\rangle|\mathbf{v}^{(b)}\rangle) \otimes |\mathbf{v}_{\text{uni}}\rangle \\ &= \frac{1}{\sqrt{2}}|\mathbf{v}^{(a)} \oplus \mathbf{v}^{(b)}\rangle \otimes |\mathbf{v}_{\text{uni}}\rangle. \end{aligned} \tag{A3}$$

The normalization factors in Eqs. (A2) and (A3) are

$$\begin{aligned} \langle \Psi_0 | \Psi_0 \rangle &= \left( \sum_{i=0}^{N-1} |v_i^{(a)}|^2 + |1 - v_i^{(a)}|^2 \right) \\ &\quad \times \left( \sum_{i=0}^{N-1} |v_i^{(b)}|^2 + |1 - v_i^{(b)}|^2 \right) \\ &= A_a A_b, \end{aligned} \tag{A4a}$$

$$\begin{aligned} \langle \Psi_1 | \Psi_1 \rangle &= \frac{1}{2} \sum_{i=0}^{N-1} (|v_i^{(a)}|^2 + |1 - v_i^{(a)}|^2 + |v_i^{(b)}|^2 \\ &\quad + |1 - v_i^{(b)}|^2) N \left( \frac{1}{2} \right)^2 \times 2 \\ &= \frac{N}{4} (A_a + A_b), \end{aligned} \tag{A4b}$$

where we have defined  $A_a = \sum_{i=0}^{N-1} |v_i^{(a)}|^2 + |1 - v_i^{(a)}|^2$  and  $A_b = \sum_{i=0}^{N-1} |v_i^{(b)}|^2 + |1 - v_i^{(b)}|^2$ . The success probability

**Algorithm 3:**  $\hat{g}_{\text{para}}(\mathbf{x}, c_0)$  (subroutine of Algorithm 4)

- 
- 1: **If**  $\mathbf{x}$  is two-dimensional:
  - 2:   prepare  $c_0$  copies of  $|\mathbf{x}\rangle$  with unitary preparation method *in parallel*
  - 3:   **Output**  $|\mathbf{x}\rangle^{\otimes c_0}$
  - 4: **Else:**
  - 5:   let  $i = \log_2[\dim(\mathbf{x})] - 1$
  - 6:   let  $\mathbf{x}^{(a)} \oplus \mathbf{x}^{(b)} = \mathbf{x}$
  - 7:   query  $g_{\text{para}}(\mathbf{x}^{(a)}, c_0)$  and  $g_{\text{para}}(\mathbf{x}^{(b)}, c_0)$  *in parallel*, get return  $|\mathbf{x}^{(a)}\rangle^{\otimes c_a}$  and  $|\mathbf{x}^{(b)}\rangle^{\otimes c_b}$
  - 8:   define  $c_{\min} = \min\{c_a, c_b\}$
  - 9:   perform transformation  $|\mathbf{x}^{(a)}\rangle \otimes |\mathbf{x}^{(b)}\rangle \rightarrow |\mathbf{x}\rangle$  for  $c_{\min}$  times *in parallel*, with  $c'(i)$  trials success
  - 10: **Output**  $|\mathbf{x}\rangle^{\otimes c'(i)}$
- 

of the projection can be calculated with the normalization factor of the quantum states before and after the projection,

$$p_+ = \frac{|\langle \Psi_1 | \Psi_1 \rangle|}{|\langle \Psi_0 | \Psi_0 \rangle|} = \frac{N A_a + A_b}{4 A_a A_b}. \tag{A5}$$

In Result 1, all amplitudes are assumed to be positive, i.e.,  $v_i^{(a,b)} \in [0, 1]$ . So we have  $A_{a,b} \in [\frac{1}{2}N, N]$ . According to Eq. (A5), the success probability satisfies

$$p_+ \in [\frac{1}{2}, 1]. \tag{A6}$$

So Result 1 holds true.

If  $v_i^{(a,b)}$  are complex values, we have  $A_{a,b} \in [\frac{1}{2}N, 5N]$ , which gives

$$p_+ \in [\frac{1}{10}, 1]. \tag{A7}$$

So  $p_+$  is lower bounded by 1/10.

**APPENDIX B: RUNTIME FOR PARALLEL PREPARATION OF A POSITIVE LABEL ENCODING STATE**

In parallel quantum state preparation (Algorithm 2 in the main text),  $c_0$  determines how many copies of each two-qubit state we should prepare. With larger  $c_0$ , it is less likely to have  $c = 0$  at line 9, and therefore the average runtime is lower. In Appendix B 1, we estimate the average runtime for  $c_0 = \lceil N^{\beta_q - 1} \rceil$  ( $1 \leq \beta_q < 2$ ) numerically; in Appendix B 2, we prove that when  $c_0 = \lceil K(N + N^{3/4}) \rceil$ , the average runtime is  $O((\ln N)^2)$ .

**1.  $c_0 = \lceil N^{\beta_q - 1} \rceil$**

The average runtimes are estimated numerically by simulating Algorithm 2. At each run, we initialize the number of steps as  $t_{\text{stp}} = 0$ . Whenever line 8 is reached, we update  $t_{\text{stp}} = t_{\text{stp}} + \log_2 \dim(\mathbf{x})$ . This is because the runtime for line 8 dominates the total runtime, and its circuit depth is proportional to  $\log_2 \dim(\mathbf{x})$ . Algorithm 2 is run 1000 times for  $N \in [4, 10]$ . We assume that the total average runtime increases polynomially as  $O(N^{\beta_t})$ . For each  $\beta_q$ , the time exponent  $\beta_t$  is estimated according to the slope of plot  $\log_2(t_{\text{stp}})$  versus  $\log_2(N)$ . Here,  $\overline{t_{\text{stp}}}$  is the mean of  $t_{\text{stp}}$  at the final step. For example, in Fig. 5, the slope for  $c_0 = 1$  is 1.52, so the average runtime  $T_{\text{pos}}$  can be estimated as  $O(N^{1.52})$ .

**2.  $c_0 = \lceil N + N^{3/4} \rceil$**

In the following, we show that by setting  $c_0 = \lceil N + N^{3/4} \rceil$ , the parallel preparation with Algorithm 2 has an average

runtime of  $O(n^2)$ . To facilitate the discussion, we introduce a variance of parallel preparation Algorithm 4. Obviously, the average runtime of Algorithm 4 is always lower than Algorithm 2. So we can just focus on Algorithm 4 in the following, and show that it has an average runtime of  $O(n^2)$ .

For a  $2^{n+1}$ -dimensional input vector  $\mathbf{x}$ , the average runtime of Algorithm 4 is  $O(n^2)$ . If one can show that, at line 10,  $\Pr[c'(n) > 0]$  [the probability that  $c'(n) > 0$ ] is larger than a nonzero constant (for arbitrarily large  $n$ ), the average runtime of Algorithm 4 can be bounded by  $O(n^2)$ .

To facilitate the discussion, we define  $c_{\text{bnd}}(n, i) \equiv 2^{n-i} + 2^{3(n-i)/4}$  [note that  $c_{\text{bnd}}(n, i) > 0$ ], and denote  $P_{n,i} \equiv \Pr[c'(i) > c_{\text{bnd}}(n, i)]$ ,  $P_n \equiv P_{n,n}$ . Now, one just needs to show that  $P_n$  is always lower bounded by a nonzero constant.

At line 7 of Algorithm 4, the probability of both  $c_a > c_{\text{bnd}}(n, i - 1)$  and  $c_b > c_{\text{bnd}}(n, i - 1)$  is  $P_{n,i-1}$ . So for  $c_{\min}$  at line 8, we have  $\Pr[c_{\min} > c_{\text{bnd}}(n, i - 1)] = P_{n,i-1}^2$ . Therefore,

$$\begin{aligned} P_{n,i} &\geq \Pr[c'(i) > c_{\text{bnd}}(n, i) | c_{\min} > c_{\text{bnd}}(n, i - 1)] \\ &\quad \times \Pr[c_{\min} > c_{\text{bnd}}(n, i - 1)] \\ &\geq \Pr[c'(i) > c_{\text{bnd}}(n, i) | c_{\min} > c_{\text{bnd}}(n, i - 1)] P_{n,i-1}^2 \\ &\geq \Pr[c'(i) > c_{\text{bnd}}(n, i) | c_{\min} = c_{\text{bnd}}(n, i - 1)] P_{n,i-1}^2. \end{aligned} \tag{B1}$$

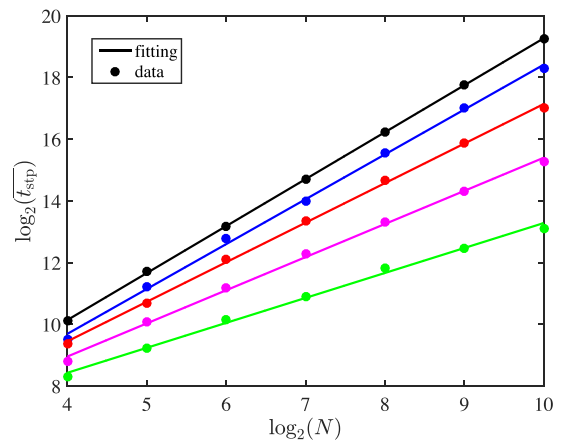


FIG. 5. Numerical results of the average runtime for Algorithm 2 for  $c_0 = 1$  (black),  $c_0 = \lceil N^{0.2} \rceil$  (blue),  $c_0 = \lceil N^{0.4} \rceil$  (red),  $c_0 = \lceil N^{0.6} \rceil$  (magenta), and  $c_0 = \lceil N^{0.8} \rceil$  (green). Dots are simulation results of  $t_{\text{stp}}$  averaged over 1000 times; lines are the corresponding linear fittings.

**Algorithm 4:**  $g_{\text{para}}(\mathbf{x}, c_0)$

- 
- 
- 1: query  $\hat{g}_{\text{para}}(\mathbf{x}, c_0)$  to obtain  $|\mathbf{x}\rangle^{\otimes c}$ .
  - 2: **If**  $c > 0$ :
  - 3:     **Output**  $|\mathbf{x}\rangle^{\otimes c}$
  - 4: **Else if**  $c = 0$ :
  - 5:     go to line 1
- 
- 

The last inequality is because  $c'(i)$  follows binomial distribution  $c'(i) \sim \mathbf{B}(c_{\min}, p_+)$  with  $p_+ > 1/2$ , and the probability that  $c'(i) > c_{\text{bnd}}(n, i)$  increases monotonically with  $c_{\min}$ .  $\Pr[c'(i) > c_{\text{bnd}}(n, i) | c_{\min} = c_{\text{bnd}}(n, i - 1)]$  is just the cumulative distribution function of binomial distribution, and according to Hoeffding's inequality, we have

$$\Pr[c'(i) > c_{\text{bnd}}(n, i) | c_{\min} = c_{\text{bnd}}(n, i - 1)] > f(n, i), \quad (\text{B2})$$

where

$$f(n, i) \equiv 1 - \exp\left(-2c_{\text{bnd}}(n, i - 1)\left(\frac{1}{2} - \frac{c_{\text{bnd}}(n, i)}{c_{\text{bnd}}(n, i - 1)}\right)^2\right). \quad (\text{B3})$$

So we have

$$P_{n,i} > P_{n,i-1}^2 f(n, i) \quad (\text{B4})$$

and

$$P_n = P_{n,n} > P_{n,n-1}^2 f(n, n) > P_{n,n-2}^2 f(n, n-1)^2 f(n) \cdots > \prod_{i=1}^{n-1} f(n, i)^{2^{n-i}}. \quad (\text{B5})$$

Because  $f(n, i) < 1$  for all  $i$ , the cumulative product  $\prod_{i=1}^{n-1} [1 - f(n, i)]^{2^{n-i}}$  decreases with  $n$  monotonically. So we have

$$P_n > \lim_{n \rightarrow \infty} P_n > \lim_{n \rightarrow \infty} \prod_{i=1}^{n-1} [1 - f(n, i)]^{2^{n-i}} > 0.006. \quad (\text{B6})$$

As  $P_n$  is lower bounded, Algorithm 4 has an average runtime of  $O(n^2)$ . Because the average runtime of Algorithm 2 is always lower than Algorithm 4, it also has an average runtime of  $O(n^2)$ .

**APPENDIX C: PROJECTION SUCCESS PROBABILITY FOR PREPARING ARBITRARY QUANTUM STATES**

Here, we discuss the projection success probability  $p_s, p'_s$ . In Appendix C1 we discuss perfect preparation and prove Result 4; in Appendix C2 we discuss cutoff preparation and prove Result 5.

**1. Perfect preparation**

With the preparation method used in Result 2 for positive vectors,  $p_s$  can be directly determined by the success probability of projecting the value qubit to  $|0\rangle$ :

$$p_s = \frac{\sum_{i=0}^{N-1} |v_i|^2}{\sum_{i=0}^{N-1} |v_i|^2 + \sum_{i=0}^{N-1} |1 - v_i|^2} \geq \frac{1}{N} \sum_{i=0}^{N-1} |v_i|^2. \quad (\text{C1})$$

$p'_s$  in Result 3 for general complex vectors is more involved. From Eqs. (5) and (6), the projection operator can be written as  $M' = |+\rangle\langle+|^{\otimes 2} \otimes \mathbb{I}_{2^{n+1}} \otimes (\mathbb{I}_{2^n} \otimes |+\rangle\langle+|^{\otimes 3})$ . The quantum state before measurement is [Eq. (5) in the main text]

$$|\Psi'_0\rangle = (|00\rangle|v^{abcd}\rangle - |01\rangle|v^{bacd}\rangle + i|10\rangle|v^{cbad}\rangle - i|11\rangle|v^{dbca}\rangle), \quad (\text{C2})$$

and the state after the projection is [Eq. (6) in the main text]

$$\begin{aligned} |\Psi'_1\rangle &= M'|\Psi_0\rangle \\ &= \frac{1}{2}|+\rangle^{\otimes 2}(|v^a\rangle - |v^b\rangle + i|v^c\rangle - i|v^d\rangle) \otimes |v^{\text{uni}}\rangle^{\otimes 3} \\ &= \frac{1}{2}|+\rangle^{\otimes 2} \sum_{i=0}^{N-1} |n, i\rangle [(v_i^{(a)} - v_i^{(b)} + iv_i^{(c)} - iv_i^{(d)})|0\rangle \\ &\quad - (v_i^{(a)} - v_i^{(b)} + iv_i^{(c)} - iv_i^{(d)})|1\rangle] \otimes |v^{\text{uni}}\rangle^{\otimes 3} \\ &= \frac{1}{2}|+\rangle^{\otimes 2} \sum_{i=0}^{N-1} v_i |n, i\rangle (|0\rangle - |1\rangle) \otimes |v^{\text{uni}}\rangle^{\otimes 3} \\ &= \frac{1}{\sqrt{2}}|+\rangle^{\otimes 2} |\psi(\mathbf{v})\rangle \otimes |-\rangle \otimes |v^{\text{uni}}\rangle^{\otimes 3}. \end{aligned} \quad (\text{C3})$$

Therefore,  $p'_s$  can be calculated as

$$p'_s = \frac{|\langle\Psi'_1|\Psi'_1\rangle|}{|\langle\Psi'_0|\Psi'_0\rangle|}. \quad (\text{C4})$$

Because  $|\langle v^a|v^a\rangle| = \sum_{i=0}^{N-1} (v_i^a)^2 + (1 - v_i^a)^2 \leq N$  and similarly for  $|v^b\rangle, |v^c\rangle, |v^d\rangle$ , we have  $|\langle v^{abcd}|v^{abcd}\rangle| \leq N^4$ . So we have  $|\langle\Psi'_0|\Psi'_0\rangle| \leq 4N^4$ . In addition, one can calculate that  $|\langle\Psi'_1|\Psi'_1\rangle| = \frac{N^3}{16} \sum_{i=0}^{N-1} |v_i|^2$ . Therefore,

$$p'_s \geq \frac{1}{64N} \sum_{i=0}^{N-1} |v_i|^2. \quad (\text{C5})$$

In the following, we prove Result 4 for sampling case 1 and sampling case 2 separately.

**a. Sampling case 1**

In case 1,  $|v_i|$  distributes uniformly in  $[0, 1]$ , so we have

$$\text{mean}\left(\frac{1}{N} \sum_{i=0}^{N-1} |v_i|^2\right) = 1/3, \quad (\text{C6a})$$

$$\text{mean}\left(\frac{1}{N} \sum_{i=0}^{N-1} |v_i|^2 + |1 - v_i|^2\right) = 2/3. \quad (\text{C6b})$$



For  $x, y > 0$ , we have

$$\begin{aligned} \Pr[p_s \geq x/y] &\geq \Pr\left[\frac{\frac{1}{N} \sum_{i=0}^{N-1} |v_i|^2}{\frac{1}{N} \sum_{i=0}^{N-1} |v_i|^2 + |1 - v_i|^2} \geq x/y\right] \\ &\geq \Pr\left[\frac{1}{N} \sum_{i=0}^{N-1} |v_i|^2 \geq x \text{ and } \frac{1}{N} \sum_{i=0}^{N-1} |v_i|^2 + |1 - v_i|^2 \leq y\right] \\ &\geq 1 - \Pr\left[\frac{1}{N} \sum_{i=0}^{N-1} |v_i|^2 < x\right] - \Pr\left[\frac{1}{N} \sum_{i=0}^{N-1} |v_i|^2 + |1 - v_i|^2 > y\right], \end{aligned} \tag{C7}$$

where we have used the relation  $\Pr[A \text{ and } B] = 1 - \Pr[A \text{ and } \bar{B}] - \Pr[\bar{A} \text{ and } B] - \Pr[\bar{A} \text{ and } \bar{B}] = 1 - \Pr[A \text{ and } \bar{B}] - \Pr[\bar{A}] \geq 1 - \Pr[\bar{B}] - \Pr[\bar{A}]$ .

According to the Chernoff bound, for any  $t_1, t_2 > 0$ , we have

$$\Pr\left[\frac{1}{N} \sum_{i=0}^{N-1} |v_i|^2 < x\right] \leq [\text{mean}(e^{-t_1|v_i|^2})]^N e^{t_1 N x} = \left[\frac{\sqrt{\pi} \text{erf}(\sqrt{t_1})}{2\sqrt{t_1}} e^{t_1 x}\right]^N \leq \left[\frac{\sqrt{\pi}}{2} \frac{e^{t_1 x}}{\sqrt{t_1}}\right]^N, \tag{C8a}$$

$$\Pr\left[\frac{1}{N} \sum_{i=0}^{N-1} |v_i|^2 + |1 - v_i|^2 > y\right] \leq \text{mean}(e^{t_2(|v_i|^2 + |1 - v_i|^2)})^N e^{-N t_2 y} = \left[\frac{e^{t_2/2} \sqrt{\pi/2} \text{erfi}(\sqrt{t_2/2})}{\sqrt{t_2}} e^{-t_2 y}\right]^N \leq \left[0.9 \frac{e^{t_2(1-y)}}{\sqrt{t_2}}\right]^N. \tag{C8b}$$

Letting  $x = \frac{1}{5}(\delta/2)^{2/N}$ ,  $y = 1 - \frac{1}{5}(\delta/2)^{2/N}$ , and setting  $t_1 = 1/(2x)$ ,  $t_2 = 1/(2 - 2y)$ , Eq. (C8) becomes

$$\Pr\left[\frac{1}{N} \sum_{i=0}^{N-1} |v_i|^2 < \frac{1}{5}(\delta/2)^{2/N}\right] \leq \delta/2, \tag{C9a}$$

$$\Pr\left[\frac{1}{N} \sum_{i=0}^{N-1} |v_i|^2 + |1 - v_i|^2 > 1 - \frac{1}{5}(\delta/2)^{2/N}\right] \leq \delta/2, \tag{C9b}$$

so Eq. (C7) becomes

$$\Pr\left[p_s \geq \frac{\frac{1}{5}(\delta/2)^{2/N}}{1 - \frac{1}{5}(\delta/2)^{2/N}}\right] \geq 1 - \delta. \tag{C10}$$

In other words,  $p_s$  is at the order of  $\Omega(\delta^{1/N})$ . For  $p'_s$ , the proof is similar.

**b. Sampling case 2**

Because  $v_i = u_i / \max(|u_i|)$ , we can rewrite Eqs. (C1) and (C5) as

$$p_s \geq \frac{1}{N} \frac{\sum_{i=0}^{N-1} |a_i|^2}{\max |a_i|^2}, \tag{C11a}$$

$$p'_s \geq \frac{1}{64N} \frac{\sum_{i=0}^{N-1} |a_i|^2}{\max |a_i|^2}. \tag{C11b}$$

In case 2,  $a_i = a_i^{(r)} + i a_i^{(m)}$ , and the real and imaginary parts of  $a_i$  are independently sampled from standard normal distribution.  $|a_i|$  follows Rayleigh distribution, and for  $a_m \in (0, \infty)$ , we have

$$\Pr[|a_i| < a_m] = 1 - e^{-a_m^2/2}. \tag{C12}$$

So the maximum over all  $|a_i|$  is given by

$$\begin{aligned} \Pr[\max |a_i|^2 < a_m^2] &= (1 - e^{-a_m^2/2})^N \\ &\geq 1 - N e^{-a_m^2/2}. \end{aligned} \tag{C13}$$

By setting  $a_m = \sqrt{2 \ln(2N/\delta)}$ , we have

$$\Pr[\max |a_i|^2 \geq 2 \ln(2N/\delta)] < \delta/2. \tag{C14}$$

Moreover, it can be calculated that

$$\text{mean}(|a_i|^2) = 2, \tag{C15a}$$

$$\text{mean}(e^{-t|a_i|^2}) = 1/(1 + 2t). \tag{C15b}$$

According to the Chernoff bound, for any  $t > 0$ , we have

$$\Pr\left[\frac{1}{N} \sum_{i=0}^{N-1} |a_i|^2 \leq x\right] \leq \left(\frac{e^{tx}}{1 + 2t}\right)^N \leq \left(\frac{e^{tx}}{2t}\right)^N. \tag{C16}$$

Letting  $x = \frac{1}{2}(\delta/2)^{1/N}$  and  $t = 1/x$ , we have

$$\Pr\left[\frac{1}{N} \sum_{i=0}^{N-1} |a_i|^2 \leq \frac{(\delta/2)^{1/N}}{2}\right] \leq \delta/2. \tag{C17}$$

Combining Eqs. (C11), (C14), and (C17), we have

$$\begin{aligned} \Pr\left[p_s \geq \frac{(\delta/2)^{1/N}}{4 \ln(2N/\delta)}\right] &\geq \Pr\left[\frac{\sum_{i=0}^{N-1} |a_i|^2}{N \max |a_i|^2} \geq \frac{(\delta/2)^{1/N}}{4 \ln(2N/\delta)}\right] \\ &\geq 1 - \Pr\left[\frac{1}{N} \sum_{i=0}^{N-1} |a_i|^2 < \frac{(\delta/2)^{1/N}}{2}\right] \\ &\quad - \Pr[\max |a_i|^2 > 2 \ln 2N/\delta] \\ &\geq 1 - \delta/2 - \delta/2 \\ &= 1 - \delta. \end{aligned} \tag{C18}$$

In other words,  $p_s = \Omega(\frac{\delta^{1/N}}{\ln(N/\delta)})$ . For  $p'_s$  the proof is similar.

**2. Cutoff preparation**

With the cutoff value  $u_{\text{cut}}$  and defining  $\tilde{v}_i \equiv \arg(u_i) \min(|u_i|/u_{\text{cut}}, 1)$ , the normalized cutoff target state becomes  $|\psi(\tilde{\mathbf{v}})\rangle = \frac{1}{\sqrt{\sum_{i=0}^{N-1} |\tilde{v}_i|^2}} \sum_{i=0}^{N-1} \tilde{v}_i |n, i\rangle$ . For sampling

case 2, it can be further rewritten as

$$|\psi(\tilde{\mathbf{v}})\rangle = \frac{1}{\|\tilde{\mathbf{a}}\|_2} \sum_{i=0}^{N-1} \tilde{a}_i |n, i\rangle, \tag{C19}$$

where  $\|\tilde{\mathbf{a}}\|_2 \equiv \sqrt{\sum_{i=0}^{N-1} |\tilde{a}_i|^2}$ , and  $\tilde{a}_i \equiv \arg(a_i) \min(|a_i|, u_{\text{cut}} \|\mathbf{a}\|_2)$ . The proof of Result 5 follows from two lemmata as follows.

*Lemma 1.* By setting  $u_{\text{cut}}^2 = \frac{8}{N} (4/\delta)^{1/N} \ln(12/(\varepsilon_{\text{th}}\delta))$ , we have

$$\Pr[F \geq 1 - \varepsilon_{\text{th}}] \geq 1 - \delta/2. \tag{C20}$$

*Proof.* Because

$$\begin{aligned} |\langle \psi(\mathbf{u}) | \psi(\tilde{\mathbf{v}}) \rangle| &= \frac{1}{\|\mathbf{a}\|_2 \|\tilde{\mathbf{a}}\|_2} \sum_i |a_i \tilde{a}_i| \\ &\geq \frac{1}{\|\mathbf{a}\|_2^2} \sum_i |a_i \tilde{a}_i| \\ &= \frac{1}{\|\mathbf{a}\|_2^2} \sum_i (|a_i|^2 - |a_i| \max(0, |a_i| - u_{\text{cut}} \|\mathbf{a}\|_2)) \\ &= 1 - \frac{1}{\|\mathbf{a}\|_2^2} \sum_i |a_i| \max(0, |a_i| - u_{\text{cut}} \|\mathbf{a}\|_2), \end{aligned} \tag{C21}$$

we have

$$\begin{aligned} F &= |\langle \psi(\mathbf{u}) | \psi(\tilde{\mathbf{v}}) \rangle|^2 \\ &\geq 1 - 2 \frac{1}{\|\mathbf{a}\|_2^2} \sum_i |a_i| \max(0, |a_i| - u_{\text{cut}} \|\mathbf{a}\|_2). \end{aligned} \tag{C22}$$

For any  $x > 0$ , we have

$$\begin{aligned} \Pr[F \geq 1 - \varepsilon_{\text{th}}] &\geq \Pr\left[\frac{1}{\|\mathbf{a}\|_2^2} \sum_i |a_i| \max(0, |a_i| - u_{\text{cut}} \|\mathbf{a}\|_2) \leq \frac{\varepsilon_{\text{th}}}{2}\right] \\ &\geq \Pr\left[\|\mathbf{a}\|_2^2 \geq Nx^2 \text{ and } \sum_i |a_i| \max(0, |a_i| - u_{\text{cut}} \|\mathbf{a}\|_2) \leq \frac{\varepsilon_{\text{th}} Nx^2}{2}\right] \\ &\geq \Pr\left[\|\mathbf{a}\|_2^2 \geq Nx^2 \text{ and } \sum_i |a_i| \max(0, |a_i| - u_{\text{cut}} \sqrt{Nx}) \leq \frac{\varepsilon_{\text{th}} Nx^2}{2}\right] \\ &\geq 1 - \Pr[\|\mathbf{a}\|_2^2 < Nx^2] - \Pr\left[\sum_i |a_i| \max(0, |a_i| - u_{\text{cut}} \sqrt{Nx}) > \frac{\varepsilon_{\text{th}} Nx^2}{2}\right] \\ &= 1 - \Pr[\|\mathbf{a}\|_2^2 < Nx^2] - \Pr\left[\sum_i \Delta_i > \frac{\varepsilon_{\text{th}} Nx^2}{2}\right], \end{aligned} \tag{C23}$$

where we have defined  $\Delta_i = |a_i| \max(0, |a_i| - u_{\text{cut}} \sqrt{Nx})$ . After some calculation, we find that for any  $t > 0$ , we have

$$\text{mean}(e^{-t|a_i|^2}) = 1/(1 + 2t) \tag{C24}$$

and

$$\text{mean}(\Delta_i) \leq 2e^{-N(u_{\text{cut}}x)^2/2}. \tag{C25}$$

From the Chernoff bound, we have

$$\Pr[\|\mathbf{a}\|_2^2 < Nx^2] \leq \left[\frac{e^{tx^2}}{1 + 2t}\right]^N. \tag{C26}$$

Letting  $x^2 = \frac{(\delta/4)^{1/N}}{2}$  and  $t = (\delta/4)^{-1/N}$ , we have

$$\Pr\left[\|\mathbf{a}\|_2^2 < N \frac{(\delta/4)^{1/N}}{2}\right] \leq \left(\frac{e^{0.5}}{1 + 2(\delta/4)^{-1/N}}\right)^N \leq \left(\frac{e^{0.5}}{2}\right)^N \left(\frac{1}{(\delta/4)^{-1/N}}\right)^N \leq \delta/4. \tag{C27}$$

Letting  $u_{\text{cut}}^2 = \frac{1}{x^2} \frac{4}{N} \ln(\frac{12}{\epsilon_{\text{th}}\delta}) = \frac{8}{N} (\frac{4}{\delta})^{1/N} \ln(\frac{12}{\epsilon_{\text{th}}\delta})$ , from the Markov inequality, we have

$$\Pr\left[\sum_i \Delta_i > N \frac{\epsilon_{\text{th}} x^2}{2}\right] \leq \frac{N \text{mean}(\Delta_i)}{N \epsilon_{\text{th}} x^2 / 2} \leq \frac{4 \exp[-\ln(144/\epsilon_{\text{th}}^2 \delta^2)]}{\epsilon_{\text{th}} x^2} \leq \frac{4 \epsilon_{\text{th}}^2 \delta^2}{144 \epsilon_{\text{th}} (\delta/4)^{1/N}} \leq \frac{\delta^2}{18(\delta/4)} \leq \delta/4. \tag{C28}$$

Combining Eqs. (C23), (C27), and (C28), Lemma 1 holds true.

Lemma 2. By setting  $u_{\text{cut}}^2 = \frac{8}{N} (4/\delta)^{1/N} \ln(12/(\epsilon_{\text{th}}\delta))$ , we have

$$\Pr[p_s \geq C_p] \geq 1 - \delta/2, \tag{C29}$$

where

$$C_p = \frac{(\delta/4)^{1/N}}{u_{\text{cut}}^2 12N \ln(4/\delta)} \tag{C30}$$

$$= \frac{(\delta/4)^{2/N}}{96 \ln(8/(\epsilon_{\text{th}}\delta)) \ln(4/\delta)}. \tag{C31}$$

Proof. According to Eq. (C1), we have  $p_s \geq \frac{1}{N} \sum_{i=0}^{N-1} \min(\frac{|a_i|^2}{(u_{\text{cut}} \|\mathbf{a}\|_2)^2}, 1)$ . So for any  $y > 0$ ,

$$\begin{aligned} \Pr\left[p_s \geq \frac{(\delta/4)^{1/N}}{3u_{\text{cut}}^2 y^2}\right] &\geq \Pr\left[\frac{1}{N} \sum_{i=0}^{N-1} \min\left(\frac{|a_i|^2}{(u_{\text{cut}} \|\mathbf{a}\|_2)^2}, 1\right) \geq \frac{(\delta/4)^{1/N}}{3u_{\text{cut}}^2 y^2}\right] \\ &\geq 1 - \Pr[\|\mathbf{a}\|_2^2 > y^2] - \Pr\left[\frac{1}{N} \sum_{i=0}^{N-1} \min\left(\frac{|a_i|^2}{(u_{\text{cut}} y)^2}, 1\right) < \frac{(\delta/4)^{1/N}}{3u_{\text{cut}}^2 y^2}\right] \\ &\geq 1 - \Pr[\|\mathbf{a}\|_2^2 > y^2] - \Pr\left[\frac{1}{N} \sum_{i=0}^{N-1} \min(|a_i|^2, (u_{\text{cut}} y)^2) < \frac{(\delta/4)^{1/N}}{3}\right] \\ &\geq 1 - \Pr[\|\mathbf{a}\|_2^2 > y^2] - \Pr\left[\sum_{i=0}^{N-1} \Delta'_i < \frac{N(\delta/4)^{1/N}}{3}\right], \end{aligned} \tag{C32}$$

where we have defined  $\Delta'_i = \min(|a_i|^2, (u_{\text{cut}} y)^2)$ . Here,  $u_{\text{cut}}^2 = \frac{8}{N} (4/\delta)^{1/N} \ln(12/(\epsilon_{\text{th}}\delta))$  as set in Lemma 1. For  $0 < t_1 < 1/2$ , we have

$$\text{mean}(e^{t_1 |a_i|^2}) = 1/(1 - 2t_1). \tag{C33}$$

So according to the Chernoff bound, we have

$$\begin{aligned} \Pr[\|\mathbf{a}\|_2^2 > y^2] &\leq e^{-t_1 y^2} [\text{mean}(e^{t_1 |a_i|^2})]^N \\ &\leq e^{-t_1 y^2} / (1 - 2t_1)^N. \end{aligned} \tag{C34}$$

Let  $y^2 = 4N \ln(8/\delta)$  and  $t_1 = 1/(4N)$ . It can be verified that

$$\begin{aligned} \Pr[\|\mathbf{a}\|_2^2 > y^2] &\leq \frac{e^{-t_1 4N \ln(8/\delta)}}{(1 - 2t_1)^N} \\ &\leq \delta/8 \left[\frac{1}{1 - 1/(2N)}\right]^N \leq \delta/4. \end{aligned} \tag{C35}$$

Moreover, we have

$$\text{mean}(e^{-t_2 \Delta'_i}) = \frac{1 + 2t_2 e^{-(u_{\text{cut}} y)^2 (1+2t_2)/2}}{1 + 2t_2}, \tag{C36a}$$

$$\text{mean}\left(\sum_{i=0}^{N-1} \Delta'_i\right) = N(2 - 2e^{-(u_{\text{cut}} y)^2 / 2}) > \frac{N(\delta/4)^{1/N}}{3}. \tag{C36b}$$

Therefore, according to the Chernoff bound,

$$\begin{aligned} \Pr\left[\sum_{i=0}^{N-1} \Delta'_i < \frac{N(\delta/4)^{1/N}}{3}\right] &\leq \left[\frac{1 + 2t_2 e^{-(u_{\text{cut}} y)^2 (1+2t_2)/2}}{1 + 2t_2}\right]^N e^{t_2 \frac{N(\delta/4)^{1/N}}{3}}. \end{aligned} \tag{C37}$$

Because  $u_{\text{cut}}^2 y^2 = 32(4/\delta)^{1/N} \ln(12/(\epsilon_{\text{th}}\delta)) \ln(8/\delta) > 32 \ln 12 \ln 8$ , when  $t_2 > 1$  we have  $2t_2 e^{-(u_{\text{cut}} y)^2 (1+2t_2)/2} < 1$ , and

$$\begin{aligned} \Pr\left[\sum_{i=0}^{N-1} \Delta'_i < \frac{N(\delta/4)^{1/N}}{3}\right] &\leq \left[\frac{1 + 1}{2t_2}\right]^N e^{t_2 \frac{N(\delta/4)^{1/N}}{3}} = \left(\frac{e^{t_2 (\delta/4)^{1/N}}}{t_2}\right)^N. \end{aligned} \tag{C38}$$

Letting  $t_2 = 3(4/\delta)^{1/N}$ , we have

$$\Pr\left[\sum_{i=0}^{N-1} \Delta'_i < N(\delta/4)^{1/N}\right] \leq \left[\frac{e}{3} \frac{1}{(4/\delta)^{1/N}}\right]^N \leq \delta/4. \tag{C39}$$

Combining Eqs. (C32), (C35), (C39), and  $y^2 = 4N \ln(8/\delta)$ , Lemma 2 holds true.

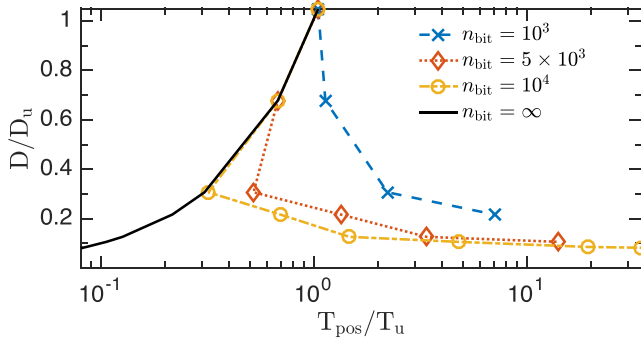


FIG. 6. Tradeoff performance between parallel low-depth preparation and the unitary preparation method.  $n_{\text{bit}}$  is the total number of available qubits.  $T_u$  and  $D_u$  are the leading order of runtime and circuit depth of the unitary preparation method in Ref. [5].

Result 5 follows directly from Lemmas 1 and 2:

$$\begin{aligned} \Pr[p_s \geq C_p \text{ and } F \geq 1 - \varepsilon_{\text{th}}] \\ \geq 1 - \Pr[p_s < C_p] - \Pr[F < 1 - \varepsilon_{\text{th}}] \\ \geq 1 - \delta. \end{aligned} \quad (\text{C40})$$

Note that

$$C_p = \Omega\left(\frac{\delta^{2/N}}{\ln \delta^{-1} \ln(\delta^{-1} \varepsilon_{\text{th}}^{-1})}\right). \quad (\text{C41})$$

For  $p'_s$ , the proof of Result 5 is similar.

#### APPENDIX D: TRADEOFF WITH UNITARY PREPARATION

As described in Algorithm 5, the tradeoff preparation method is similar to Algorithm 2 in the main text except that at line 2, unitary preparation is applied when the input state is  $2^{n_u}$ -dimensional. Algorithm 2 can be considered as an extreme case of Algorithm 5 when  $n_u = 1$ . For larger  $n_u$ , the circuit depth is higher but the average runtime is lower.

As an example, in Fig. 6 we consider the preparation of a 16-qubit target state. We use the case 2 sampling method, and the results are averaged over 100 random states. When estimating circuit depth and runtime, all controlled-swap gates and control-control swap gates are decomposed into single-qubit and CNOT gates. As can be seen, with more available

ancillary qubits we have both a lower circuit depth (of single-qubit and CNOT gates) and a lower runtime.

#### APPENDIX E: LOWER BOUND OF THE CIRCUIT DEPTH FOR PREPARING AN ARBITRARY QUANTUM STATE

Our protocols show that the state preparation can be done with  $O(n^2)$  circuit depth. The question remains whether there is a lower bound of the circuit depth for state preparation, and how close our protocol is to the fundamental limit. In this Appendix, we will address this problem under the following framework:

- (i) Initially,  $\mathbf{v}$  is stored classically and there is no other prior knowledge about the target state.
- (ii) All operations are applied on at most  $k$  qubits.
- (iii) Ancillary qubits are allowed, and qubits have all-to-all connections.

Under the above framework, we have Result 6. The main idea of the proof for Result 6 is as follows. First, we introduce the concept of a light cone, and we show that the light cone size required for quantum state preparation is  $\Omega(N)$ , as there are a total of  $N$  elements in  $\mathbf{v}$ . Second, we show that in order to obtain a light cone with  $\Omega(N)$ , at least  $\Omega(n)$  layers of local operations are required.

We begin by introducing several definitions.

*Qubit connections and light cone.* At each layer of the quantum circuit, we make a grouping of all qubits. We denote  $\pi_m^{(i)}$  as the  $i$ th group of qubits at the  $m$ th layer. Operations are applied only among the qubits in the same group. For example, if there are four qubits ( $n_{\text{tot}} = 4$ ), and  $\pi_1^{(1)} = \{1, 4\}$ ,  $\pi_1^{(2)} = \{2, 3\}$ , it means that at the first layer, qubits with label 1,4 are connected to each other, and qubits with label 2,3 are connected to each other. We also denote the group containing qubit  $j$  at the  $l$ th layer as  $\pi'_l(j)$ . In the example above, we have  $\pi'_1(1) = \{1, 4\}$ ,  $\pi'_1(2) = \{2, 3\}$ ,  $\pi'_1(3) = \{2, 3\}$ , and  $\pi'_1(4) = \{1, 4\}$ . Note that at each layer, each qubit belongs to only *one* group.

The light cone of qubit  $j$  is defined as all qubits having a connection to it. More rigorously, for an  $L$ -layer circuit, the light cone of qubit  $j$  is

$$S(j) \equiv \mathcal{S}_1(j), \quad (\text{E1})$$

---

#### Algorithm 5: $f_{\text{tradeoff}}(\mathbf{x}, c_0, n_u)$

---

- 1: **If**  $\mathbf{x}$  is  $2^{n_u}$ -dimensional:
  - 2:   prepare  $c_0$  copies of  $|\mathbf{x}\rangle$  with unitary preparation method in parallel
  - 3:   **Output**  $|\mathbf{x}\rangle^{\otimes c_0}$
  - 4: **Else:**
  - 5:   let  $\mathbf{x}^{(a)} \oplus \mathbf{x}^{(b)} = \mathbf{x}$
  - 6:   query  $f_{\text{tradeoff}}(\mathbf{x}^{(a)}, c_0, n_u)$  and  $f_{\text{tradeoff}}(\mathbf{x}^{(b)}, c_0, n_u)$  in parallel, get return  $|\mathbf{x}^{(a)}\rangle^{\otimes c_a}$  and  $|\mathbf{x}^{(b)}\rangle^{\otimes c_b}$
  - 7:   define  $c_{\text{min}} = \min\{c_a, c_b\}$
  - 8:   perform transformation  $|\mathbf{x}^{(a)}\rangle \otimes |\mathbf{x}^{(b)}\rangle \rightarrow |\mathbf{x}\rangle$  for  $c_{\text{min}}$  times in parallel, with  $c$  trials success
  - 9:   **If**  $c = 0$ :
  - 10:    go to line 6
  - 11:   **Else:**
  - 12:    **Output**  $|\mathbf{x}\rangle^{\otimes c}$
-

where

$$\mathcal{S}_m(j) \equiv \begin{cases} \pi'_m(j), & m = L, \\ \bigcup_{k \in \mathcal{S}_{m+1}(j)} \pi'_m(k), & 1 \leq m < L. \end{cases} \quad (\text{E2})$$

Because the operations are applied on at most  $k$  qubits, the light cone size [number of elements of  $\mathcal{S}(j)$ ] satisfies  $|\mathcal{S}(j)| \leq k^L$ .

*Quantum operations and classical encoding of  $\mathbf{v}$ .* If there are a total of  $N_{\text{tot}}$  qubits ( $N_{\text{tot}}$  is even), the operations at the  $m$ th layer generally take the following form:

$$\mathcal{E}_m = \mathcal{E}_{\pi_m^{(N_{\text{tot}}/2)}} \circ \dots \circ \mathcal{E}_{\pi_m^{(2)}} \circ \mathcal{E}_{\pi_m^{(1)}}, \quad (\text{E3})$$

where  $\mathcal{E}_{\pi_m^{(i)}}$  can be arbitrary quantum operations (with unitary, measurement, and postselection) in the subspace containing qubits in  $\pi_m^{(i)}$ . The full operation of an  $L$ -layer quantum circuit is just

$$\mathcal{E} = \mathcal{E}_L \circ \dots \circ \mathcal{E}_2 \circ \mathcal{E}_1. \quad (\text{E4})$$

For simplicity, we restrict ourselves to  $\mathbf{v} \in [0, 1]^N$ , and the generalization to  $\mathbf{v} \in \mathbb{C}$  is straightforward. As  $\mathbf{v}$  has a total of  $N$  elements, it takes at least  $N$  qubits (or a classical bit) to store  $\mathbf{v}$ . Without a loss of generality, we assume that  $\mathbf{v}$  is stored in the following form:

$$\bigotimes_{i=0}^{N-1} (|v_i\rangle\langle v_i|)^{\otimes N_c}, \quad (\text{E5})$$

where we have allowed multiple copies ( $N_c$ ) of the classical information. We also assume that the  $n$ -qubit system encoding  $\mathbf{v}$  is initialized to  $\rho_{\text{enc}}$ , and there is an ancillary qubits system initialized to  $\rho_{\text{anc}}$ .

The target state takes the form of Eq. (1) in the main text, so in the most general cases, quantum state preparation

processes,  $\mathcal{E}$ , should satisfy the following:

$$\mathcal{E} \left( \rho_{\text{enc}} \otimes \rho_{\text{anc}} \otimes \bigotimes_{i=0}^{N-1} (|v_i\rangle\langle v_i|)^{\otimes N_c} \right) = \alpha(\mathbf{v}) |\psi(\mathbf{v})\rangle\langle\psi(\mathbf{v})| \otimes \rho' \quad (\text{E6})$$

for all  $\mathbf{v} \in [0, 1]^N$ . Here  $\alpha(\mathbf{v}) \in (0, 1]$  as postselections are allowed, and  $\rho'$  can be an arbitrary quantum state of the joint system of ancillary qubits and the qubits encoding the classical information. Our goal is to find the lower bound of  $L$  in Eq. (E4) for quantum operations satisfying Eq. (E6).

We first introduce a Lemma relating the light cone to the reduced density matrix (RDM) of the final output states. Lemma 3 can be considered as a generalization of Lemma 2.3 in [55].

*Lemma 3.* The RDM of qubit  $j$  of the final output state depends only on the RDM of the input state for the subsystem containing qubits in its light cone  $\mathcal{S}(j)$ .

*Proof.* The RDM of the final output state for qubit  $j$  depends only on the RDM of  $\mathcal{S}_L(j)$  at the  $(L - 1)$ th layer.

Similarly, the RDM of  $\mathcal{S}_m(j)$  at the  $m$ th layer depends only on the RDM of  $\mathcal{S}_{m-1}(j)$  at the  $(m - 1)$ th layer. Therefore, the RDM of qubit  $j$  of the final output state depends only on the RDM of the total input state of  $\mathcal{S}_1(j) = \mathcal{S}(j)$ , i.e., the light cone of qubit  $j$ .

The proof of Result 6 follows directly from Lemma 3 as follows:

*Proof.* According to Eq. (E6), the final state of the encoding system and the RDM of each qubit in it depends on at least one copy of the state  $\bigotimes_{i=0}^{N-1} |v_i\rangle\langle v_i|$ .

According to Lemma 3, for qubits in the encoding system, the light cone  $\mathcal{S}(j)$  contains at least  $N$  qubits, i.e.,  $|\mathcal{S}(j)| \geq 2^n$ . If the elementary quantum operations are  $k$ -local, for an  $L$ -layer quantum circuit we have  $|\mathcal{S}(j)| \leq k^L$ . Therefore, in order to obtain an operation satisfying Eq. (E6), the circuit depth is lower bounded by  $L \geq \frac{\ln 2}{\ln k} n = O(n)$ .

[1] M. A. Nielsen and I. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, 2000).  
 [2] P. Kaye and M. Mosca, Quantum networks for generating arbitrary quantum states, in *Optical Fiber Communication Conference and International Conference on Quantum Information* (Optical Society of America, Washington, DC, 2001).  
 [3] L. Grover and T. Rudolph, Creating superpositions that correspond to efficiently integrable probability distributions, [arXiv:quant-ph/0208112](https://arxiv.org/abs/quant-ph/0208112).  
 [4] M. Möttönen, J. J. Vartiainen, V. Bergholm, and M. M. Salomaa, Transformation of quantum states using uniformly controlled rotations, *Quantum. Inf. Comput.* **5**, 467 (2005).  
 [5] M. Plesch and Č. Brukner, Quantum-state preparation with universal gate decompositions, *Phys. Rev. A* **83**, 032302 (2011).  
 [6] M.-H. Yung, J. D. Whitfield, S. Boixo, D. G. Tempel, and A. Aspuru-Guzik, Introduction to quantum algorithms for physics and chemistry, *Adv. Chem. Phys.* **67** (2014).

[7] R. Iten, R. Colbeck, I. Kukuljan, J. Home, and M. Christandl, Quantum circuits for isometries, *Phys. Rev. A* **93**, 032318 (2016).  
 [8] J. Zhao, Y.-C. Wu, G.-C. Guo, and G.-P. Guo, State preparation based on quantum phase estimation, [arXiv:1912.05335](https://arxiv.org/abs/1912.05335).  
 [9] A. W. Harrow, A. Hassidim, and S. Lloyd, Quantum Algorithm for Linear Systems of Equations, *Phys. Rev. Lett.* **103**, 150502 (2009).  
 [10] L. Wossnig, Z. Zhao, and A. Prakash, Quantum Linear System Algorithm for Dense Matrices, *Phys. Rev. Lett.* **120**, 050502 (2018).  
 [11] N. Wiebe, D. Braun, and S. Lloyd, Quantum Algorithm for Data Fitting, *Phys. Rev. Lett.* **109**, 050505 (2012).  
 [12] S. Lloyd, M. Mohseni, and P. Rebentrost, Quantum principal component analysis, *Nat. Phys.* **10**, 631 (2014).  
 [13] P. Rebentrost, M. Mohseni, and S. Lloyd, Quantum Support Vector Machine for Big Data Classification, *Phys. Rev. Lett.* **113**, 130503 (2014).

- [14] S. Lloyd, Universal quantum simulators, *Science* **273**, 1073 (1996).
- [15] D. Aharonov and A. Ta-Shma, Adiabatic quantum state generation and statistical zero knowledge, in *Proceedings of the Thirty-fifth Annual ACM Symposium on Theory of Computing* (Association for Computing Machinery, New York, 2003), pp. 20–29.
- [16] G. H. Low and I. L. Chuang, Optimal Hamiltonian Simulation by Quantum Signal Processing, *Phys. Rev. Lett.* **118**, 010501 (2017).
- [17] J. Biamonte, P. Wittek, N. Pancotti, P. Rebentrost, N. Wiebe, and S. Lloyd, Quantum machine learning, *Nature (London)* **549**, 195 (2017).
- [18] K. H. Wan, O. Dahlsten, H. Kristjánsson, R. Gardner, and M. Kim, Quantum generalisation of feedforward neural networks, *npj Quantum Inf.* **3**, 36 (2017).
- [19] K. Beer, D. Bondarenko, T. Farrelly, T. J. Osborne, R. Salzmann, D. Scheiermann, and R. Wolf, Training deep quantum neural networks, *Nat. Commun.* **11**, 808 (2020).
- [20] J. Romero, J. P. Olson, and A. Aspuru-Guzik, Quantum autoencoders for efficient compression of quantum data, *Quantum Sci. Technol.* **2**, 045001 (2017).
- [21] D. Bondarenko and P. Feldmann, Quantum Autoencoders to Denoise Quantum Data, *Phys. Rev. Lett.* **124**, 130502 (2020).
- [22] X. Wang, Y. Ma, M.-H. Hsieh, and M.-H. Yung, Quantum speedup in adaptive boosting of binary classification, *Sci. China Phys. Mech.* **64**, 220311 (2021).
- [23] X.-M. Zhang, W. Kong, M. U. Farooq, M.-H. Yung, G. Guo, and X. Wang, Generic detection-based error mitigation using quantum autoencoders, *Phys. Rev. A* **103**, L040403 (2021).
- [24] In pp. 198–200, Sec. 4.5.4 of Ref. [1], it is proven that with constant-weight quantum gates and no ancillary qubit, there are states of  $n$  qubits that take  $\Omega(N/\log n)$  constant-weight operations to approximate to within a fixed distance (Eq. 4.85).
- [25] J. A. Cortese and T. M. Braje, Loading classical data into a quantum computer, [arXiv:1803.01958](https://arxiv.org/abs/1803.01958).
- [26] I. F. Araujo, D. K. Park, F. Petruccione, and A. J. da Silva, A divide-and-conquer algorithm for quantum state preparation, *Sci. Rep.* **11**, 6329 (2021).
- [27] S. Johri, S. Debnath, A. Mocherla, A. Singk, A. Prakash, J. Kim, and I. Kerenidis, Nearest centroid classification on a trapped ion quantum computer, *npj Quantum Inf.* **7**, 122 (2021).
- [28] A. Prakash, Quantum algorithms for linear algebra and machine learning., Ph.D. thesis, UC Berkeley (2014).
- [29] A. N. Soklakov and R. Schack, Efficient state preparation for a register of quantum bits, *Phys. Rev. A* **73**, 012307 (2006).
- [30] S. Lloyd, M. Mohseni, and P. Rebentrost, Quantum algorithms for supervised and unsupervised machine learning, [arXiv:1307.0411](https://arxiv.org/abs/1307.0411).
- [31] S. Lloyd and C. Weedbrook, Quantum Generative Adversarial Learning, *Phys. Rev. Lett.* **121**, 040502 (2018).
- [32] P.-L. Dallaire-Demers and N. Killoran, Quantum generative adversarial networks, *Phys. Rev. A* **98**, 012324 (2018).
- [33] M. Schuld and N. Killoran, Quantum Machine Learning in Feature Hilbert Spaces, *Phys. Rev. Lett.* **122**, 040504 (2019).
- [34] M. Schuld, A. Bocharov, K. M. Svore, and N. Wiebe, Circuit-centric quantum classifiers, *Phys. Rev. A* **101**, 032308 (2020).
- [35] S. Lu, L.-M. Duan, and D.-L. Deng, Quantum adversarial machine learning, *Phys. Rev. Res.* **2**, 033212 (2020).
- [36] J. Watrous, *The Theory of Quantum Information* (Cambridge University Press, Cambridge, 2018).
- [37] X. Sun, G. Tian, S. Yang, P. Yuan, and S. Zhang, Asymptotically optimal circuit depth for quantum state preparation and general unitary synthesis, [arXiv:2108.06150](https://arxiv.org/abs/2108.06150).
- [38] P. Selinger, Efficient clifford+ $T$  approximation of single-qubit operators, [arXiv:1212.6253](https://arxiv.org/abs/1212.6253).
- [39] J. Preskill, Quantum computing in the nisq era and beyond, *Quantum* **2**, 79 (2018).
- [40] F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends, R. Biswas, S. Boixo, F. G. Brandao, D. A. Buell *et al.*, Quantum supremacy using a programmable superconducting processor, *Nature (London)* **574**, 505 (2019).
- [41] G. AI Quantum *et al.*, Hartree-Fock on a superconducting qubit quantum computer, *Science* **369**, 1084 (2020).
- [42] M. Gong, S. Wang, C. Zha, M.-C. Chen, H.-L. Huang, Y. Wu, Q. Zhu, Y. Zhao, S. Li, S. Guo *et al.*, Quantum walks on a programmable two-dimensional 62-qubit superconducting processor, *Science* **372**, 948 (2021).
- [43] M. Cerezo, A. Arrasmith, R. Babbush, S. C. Benjamin, S. Endo, K. Fujii, J. R. McClean, K. Mitarai, X. Yuan, L. Cincio *et al.*, Variational quantum algorithms, *Nat. Rev. Phys.* **3**, 625 (2021).
- [44] S. Endo, Z. Cai, S. C. Benjamin, and X. Yuan, Hybrid quantum-classical algorithms and quantum error mitigation, *J. Phys. Soc. Jpn.* **90**, 032001 (2021).
- [45] X. Xu, J. Sun, S. Endo, Y. Li, S. C. Benjamin, and X. Yuan, Variational algorithms for linear algebra, *Sci. Bull.* **66**, 2181 (2021).
- [46] J. Liu, J. Sun, and X. Yuan, Towards a variational jordan-lee-preskill quantum algorithm, [arXiv:2109.05547](https://arxiv.org/abs/2109.05547).
- [47] X. Yuan, J. Sun, J. Liu, Q. Zhao, and Y. Zhou, Quantum Simulation with Hybrid Tensor Networks, *Phys. Rev. Lett.* **127**, 040501 (2021).
- [48] S. Endo, J. Sun, Y. Li, S. C. Benjamin, and X. Yuan, Variational Quantum Simulation of General Processes, *Phys. Rev. Lett.* **125**, 010501 (2020).
- [49] J. Sun, S. Endo, H. Lin, P. Hayden, V. Vedral, and X. Yuan, Perturbative quantum simulation, [arXiv:2106.05938](https://arxiv.org/abs/2106.05938).
- [50] T. Peng, A. W. Harrow, M. Ozols, and X. Wu, Simulating large quantum circuits on a small quantum computer, *Phys. Rev. Lett.* **125**, 150504 (2020).
- [51] H.-S. Zhong, H. Wang, Y.-H. Deng, M.-C. Chen, L.-C. Peng, Y.-H. Luo, J. Qin, D. Wu, X. Ding, Y. Hu *et al.*, Quantum computational advantage using photons, *Science* **370**, 1460 (2020).
- [52] M.-H. Yung, Quantum supremacy: Some fundamental concepts, *Natl. Sci. Rev.* **6**, 22 (2019).
- [53] B. Wu, B. Cheng, F. Jia, J. Zhang, M.-H. Yung, and X. Sun, Speedup in classical simulation of gaussian boson sampling, *Sci. Bull.* **65**, 832 (2020).
- [54] X. Chen, B. Cheng, Z. Li, X. Nie, N. Yu, M.-H. Yung, and X. Peng, Experimental cryptographic verification for near-term quantum cloud computing, *Sci. Bull.* **66**, 23 (2021).
- [55] B. Barak, C.-N. Chou, and X. Gao, Spoofing linear cross-entropy benchmarking in shallow quantum circuits, [arXiv:2005.02421](https://arxiv.org/abs/2005.02421).