


Quantum Wasserstein distance between unitary operations

Xinyu Qiu ^{1,*}, Lin Chen ^{1,2,†} and Li-Jun Zhao^{1,‡}

¹*LMIB (Beihang University), Ministry of Education, and School of Mathematical Sciences, Beihang University, Beijing 100191, China*

²*International Research Institute for Multidisciplinary Science, Beihang University, Beijing 100191, China*

 (Received 16 June 2023; accepted 6 June 2024; published 1 July 2024)

Quantifying the effect of noise on unitary operations is an essential task in quantum information processing. We propose the quantum Wasserstein distance between unitary operations, which shows an explanation for quantum circuit complexity and characterizes the local distinguishability of multiqubit operations. We show analytical calculations of the distance between identity and widely used quantum gates including SWAP, CNOT, and other controlled gates. As an application, we estimate the closeness between quantum gates in a circuit and show that the noisy operation simulates the ideal one well when they become close under the distance. Further, we introduce the W_1 error rate by the distance and establish the relation between the W_1 error rate and two practical cost measures of recovery operations in quantum error correction under typical noise scenarios. These applications allow the distance to quantify the effect of noise on unitary operations from the perspective of experiment resources, which cannot be achieved by existing distance.

DOI: [10.1103/PhysRevA.110.012412](https://doi.org/10.1103/PhysRevA.110.012412)

I. INTRODUCTION

Recent progress in quantum information processing has led to prominent applications, such as simulation [1], control into computation [2], and machine learning [3,4]. Real-world imperfections exist and current quantum computers are inevitably noisy. So it is essential to quantify how much noise can influence a quantum operation. It is necessary to analyze the similarity measure between ideal and actual operation in a noisy environment. Generally, the similarity measure between operations can be induced by that between quantum states. Some prominent measures for states, including trace distance [5], quantum fidelity, and relative entropy [6], are unitary invariant. The unitary invariance property makes the distance between any couple of states with orthogonal supports maximal, which is not desirable for some quantum information processing tasks, such as quantum error correction. Recently, some nonunitarily invariant Wasserstein distances of different orders has been proposed [7–9]. In particular, the quantum Wasserstein distance of order one (W_1 distance) [7] derives numerous applications, including quantum differential privacy [10] and quantum circuit complexity [11]. This paper aims to construct the similarity measure for unitary operations by the W_1 distance between states.

Various distances, or similarity measures, between operations have been proposed. Although no distance is generally suitable for all quantum information processing tasks, each kind of distance characterizes operations in a specific application. The Schatten 1-norm-induced distance is used to discriminate unitary operations [12], and the Schatten 2-norm-induced distance can be efficiently estimated in quan-

tum circuit [13]. Gate fidelity is experimentally convenient [14], but the connection with fault-tolerant computation is not direct. An alternative is the diamond norm-induced distance that yields tighter estimates of operation performance. These widely used distances are unitarily invariant and characterize global properties of quantum operations, while their local distinguishability may not be depicted. With regard to the nonunitarily invariant distance, recently the W_1 distance between channels has been analyzed by proposing the notion of “neighboring channels” [15], while its relation with quantum information processing tasks remains to be established. Following a different approach, the W_1 distance between operations can also be defined by the maximum deviation of their effects on overall quantum states. The formulation is inspired by the discrimination of unitary operations [12]. Compared with the existing distance, it characterizes the local distinguishability of operations and relates to quantum circuit complexity and quantum error correction (QKD). These properties allow it to quantify the effect of noise on unitary operations from the perspective of experiment resources, which cannot be realized by existing distances. This is the motivation of this paper.

In this paper we introduce a similarity measure of unitary operations, named the quantum Wasserstein distance between unitary operations. We show the properties and analytical calculations of the distance. Compared with the existing distance, it characterizes the local distinguishability of operations and shows explanations for circuit complexity. By these characterizing properties, two applications of the distance are presented. First, the distance is applied to estimate the closeness between two sequences of gates in a quantum circuit, where the noisy operation simulates the ideal one well when they are close under the distance. Next, the W_1 gate error rate is introduced by the distance, which quantifies the performance of quantum gates in a noisy environment from the perspective of physical resources. The relation between the W_1 error rate

*Contact author: xinyuqiu@buaa.edu.cn

†Contact author: linchen@buaa.edu.cn

‡Contact author: zhaolijun@buaa.edu.cn

and two practical cost measures of recovery operations is established, which makes the W_1 error rate a figure of merit concerning a specific gate and the experimental requirement to eliminate noise acting on the gate.

The rest of the paper is organized as follows. In Sec. II we introduce the notations, some properties of the quantum W_1 distance between states, the formalism of gate error rate, circuit cost, and experiment cost. In Sec. III we show the definition, properties and calculation of quantum W_1 distance between operations. In Sec. IV we show the estimation of the closeness between operations in quantum circuits. In Sec. V we introduce the W_1 gate error rate and show the noisy implementation of arbitrary single-qubit gate and CNOT gate under typical noise scenarios. We conclude in Sec. VI.

II. PRELIMINARIES AND NOTATIONS

In this section we show the notations and some facts used in this paper. In Sec. II A we present the notations of this paper. In Sec. II B we show the definition and some properties of quantum W_1 distance between states. In Sec. II C we introduce the definition of the average gate fidelity and gate error rate induced by different norms. In Sec. II D we introduce two practical circuit complexity measures, the circuit and experiment cost.

A. Notations

Let $\{|1\rangle, |2\rangle, \dots, |d\rangle\}$ be the canonical basis of \mathbb{C}^d , and $\mathcal{H}_n = (\mathbb{C}^d)^{\otimes n}$ be the Hilbert space of n qudits. We denote the set of quantum states on \mathcal{H}_n by \mathcal{S}_n , the set of traceless, self-adjoint linear operators on \mathcal{H}_n by \mathcal{M}_n , and the set of unitary operations acting on \mathcal{S}_n by \mathcal{U}_n .

Some well-known single-qubit gate include the Hadamard gate $H = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$ and the Pauli matrices $\sigma_x = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$, $\sigma_y = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}$, $\sigma_z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$. The two-qubit gates include

the CNOT gate $U_{CN} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}$, the controlled-

Z gate $U_{CZ} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix}$, the SWAP gate $U_{SW} =$

$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$, and the generalized controlled phase gate

$U_{CP} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & e^{i\theta} \end{bmatrix}$. The single-qudit Pauli gate $X =$

$\sum_{q=0}^3 |q \oplus 1\rangle \langle q|$ and $Z = \sum_{q=0}^3 \omega^q |q\rangle \langle q|$ for $\omega = e^{2\pi i/d}$.

The Schatten p -norm for arbitrary matrix $A \in \mathbb{C}^{N \times M}$ and $p \in [1, \infty)$ is defined as $\|A\|_p = [\text{Tr}(A^\dagger A)^{\frac{p}{2}}]^{\frac{1}{p}}$. By setting $p = 1$ into the definition of $\|\cdot\|_p$, one has the Schatten 1-norm (trace norm) given by $\|A\|_1 = \text{Tr} \sqrt{A^\dagger A}$, which is equal to the sum of singular values of A . For two states $\rho, \sigma \in \mathcal{S}_n$, $\frac{1}{2} \|\rho - \sigma\|_1$ is typically denoted as the trace distance between ρ and σ .

B. The quantum Wasserstein distance of order 1 between states

The well-known similarity measures between quantum states characterize the global distinguishability of states. For

specific applications, such as quantum error correction and quantum machine learning, it is desirable to show the local distinguishability and robustness against local perturbations on the input states. The quantum W_1 distance between states [7] is a good candidate for this task. We show the definition and some essential properties of the quantum Wasserstein distance, which will be used in this paper.

First, we show some basic definitions. The quantum Wasserstein norm of order 1 is a kind of unique norm on \mathcal{M}_n [7]. Following the quantum W_1 norm, the quantum Wasserstein distance between states ρ and σ is naturally obtained. They are defined as follows.

Definition 1. We define the quantum W_1 norm on \mathcal{M}_n as, for any $X \in \mathcal{M}_n$,

$$\|X\|_{W_1} = \frac{1}{2} \min \left(\sum_{i=1}^n \|X^{(i)}\|_1 : X^{(i)} \in \mathcal{M}_n, \right. \\ \left. \times \text{Tr}_i X^{(i)} = 0, X = \sum_{i=1}^n X^{(i)} \right), \quad (1)$$

where $\text{Tr}_i[\cdot]$ denotes the partial trace over the i th subsystem. By choosing $X = \rho - \sigma$, the quantum Wasserstein distance of order 1 between two quantum states $\rho, \sigma \in \mathcal{S}_n$ is defined as

$$W_1(\rho, \sigma) \\ = \min \left\{ \sum_{i=1}^n c_i : c_i \geq 0, \rho - \sigma = \sum_{i=1}^n c_i (\rho^{(i)} - \sigma^{(i)}), \right. \\ \left. \times \rho^{(i)}, \sigma^{(i)} \in \mathcal{S}_n, \text{Tr}_i \rho^{(i)} = \text{Tr}_i \sigma^{(i)} \right\}. \quad (2)$$

We list the properties of the available quantum W_1 distance below. They will be used for the derivation and applications of the quantum W_1 distance between operations.

The following fact shows that the quantum W_1 norm keeps the upper and lower bounds in terms of the trace norm.

Lemma 2. (relation with the trace norm, [7]) For any $X \in \mathcal{M}_n$, $\frac{1}{2} \|X\|_1 \leq \|X\|_{W_1} \leq \frac{n}{2} \|X\|_1$. Moreover, if $\text{Tr}_i X = 0$ for some $i \in [n]$, then $\|X\|_{W_1} = \frac{1}{2} \|X\|_1$. That is, for any $\rho, \sigma \in \mathcal{S}_n$ such that $\text{Tr}_i \rho = \text{Tr}_i \sigma$ for some $i \in [n]$, $\|\rho - \sigma\|_{W_1} = \frac{1}{2} \|\rho - \sigma\|_1$.

Lemma 3 and Corollary 4 show that the quantum W_1 distance is additive with respect to the tensor product and partial trace. It is a characteristic property that cannot be satisfied by the trace distance. In this paper, they are used for calculating the quantum W_1 distance between unitary operations and deriving its properties.

Lemma 3. (tensorization, [7]) For any $X \in \mathcal{M}_{m+n}$, $\|X\|_{W_1} \geq \|\text{Tr}_{m+1\dots m+n} X\|_{W_1} + \|\text{Tr}_{1\dots m} X\|_{W_1}$, and for any $\rho, \sigma \in \mathcal{S}_{m+n}$, $\|\rho - \sigma\|_{W_1} \geq \|\rho_{1\dots m} - \sigma_{1\dots m}\|_{W_1} + \|\rho_{m+1\dots m+n} - \sigma_{m+1\dots m+n}\|_{W_1}$, where $\eta_{1\dots m} = \text{Tr}_{m+1, m+2, \dots, m+n} \eta$ and $\eta_{m+1, m+2, \dots, m+n} = \text{Tr}_{1\dots m} \eta$ with $\eta \in \{\rho, \sigma\}$. Moreover, for any $\rho', \sigma' \in \mathcal{S}_m$ and $\rho'', \sigma'' \in \mathcal{S}_n$, $\|\rho' \otimes \rho'' - \sigma' \otimes \sigma''\|_{W_1} = \|\rho' - \sigma'\|_{W_1} + \|\rho'' - \sigma''\|_{W_1}$.

Corollary 4. (lower bound for W_1 distance, [7]) For any $\rho, \sigma \in \mathcal{S}_n$, $\|\rho - \sigma\|_{W_1} \geq \frac{1}{2} \sum_{i=1}^n \|\rho_i - \sigma_i\|_1$, where $\eta_i = \text{Tr}_{1, \dots, i-1, i+1, \dots, n} \eta$ with $\eta \in \{\rho, \sigma\}$, and the equality holds whenever both ρ and σ are product states.

The following observation states that the quantum W_1 distance recovers the Hamming distance for the quantum states described by a canonical basis. It contributes to the calculation of the quantum W_1 distance between unitary operations.

Lemma 5. (recovery of Hamming distance, [7]) The quantum W_1 distance between vectors of the canonical basis coincides with the Hamming distance $\| |x\rangle\langle x| - |y\rangle\langle y| \|_{W_1} = h(x, y)$, with $x, y \in [d]^n$. Here the Hamming distance between $x, y \in [d]^n$ is the number of different components, $h(x, y) = |\{i \in [n] : x_i \neq y_i\}|$ with $x = (x_1, \dots, x_n)^T$ and $y = (y_1, \dots, y_n)^T$.

C. Average gate fidelity and gate error rate

In practice, noise is inevitable during the implementation of quantum gates. It is important to quantify how much the noise affects quantum states in a system. Average gate fidelity [14] is first proposed to accomplish such a task. Suppose the ideal and actual implementation of quantum gate U is denoted by channels $\mathcal{G}_{id}(\cdot) = U(\cdot)U^\dagger$ and $\mathcal{V}(\cdot)$ and their resulting states are ρ_{id} and ρ_{ac} , respectively. Averaging over pure state input with respect to the Haar measure derives the average gate fidelity,

$$\eta := \int d\mu(\rho) \text{Tr}[(U\rho U^\dagger)\mathcal{V}(\rho)].$$

Then the gate error rate [16–18] is proposed, whose upper bound is an appropriate measure to assess progress towards fault-tolerant quantum computation. It is derived by the error rate of probability distributions for an arbitrary set of POVM [16],

$$e_1(U, \mathcal{V}) := \max_{\rho} d_1(\rho_{ac}, \rho_{id}) = \frac{1}{2} \max_{\rho} \|U\rho U^\dagger - \mathcal{V}(\rho)\|_1.$$

The above definition is amended by maximizing over the original space \mathcal{H}_n and ancillary space \mathcal{H}'_n . Besides the notations in Sec. II A, we denote by \mathcal{S}'_n and \mathcal{U}'_n the set of quantum states and operations on auxiliary space \mathcal{H}'_n . The diamond norm is then defined as $\|X\|_{\diamond} := \sup_{\mathcal{U}'_n} \sup_{\rho \in \mathcal{S}'_n \otimes \mathcal{S}_n} \|X \otimes I(\rho)\|_1$, where $X \in \mathcal{U}_n$, $I \in \mathcal{U}'_n$. Another definition of gate error rate is derived by [18],

$$e_{\diamond}(U, \mathcal{V}) = d_{\diamond}(\mathcal{G}_{id}, \mathcal{V}) := \frac{1}{2} \|U\rho U^\dagger - \mathcal{V}(\rho)\|_{\diamond}.$$

The error rate $e_{\diamond}(U, \mathcal{V})$ has a better explanation in fault-tolerance quantum computation than fidelity and $e_1(U, \mathcal{V})$. However, neither of the existing error rates shows explanations for quantum circuit complexity and QKD. To deal with it, we propose the W_1 error rate and establish its lower bounds with the help of W_1 distance between operations. It will be presented in Sec. V.

D. Circuit and experiment cost

Quantum circuit complexity of a unitary operation is defined as the minimal number of basic gates required to synthesize a desired unitary operation [19]. Computing the circuit complexity of unitaries is challenging, and no efficient algorithms are known [20]. It has been proved that the circuit complexity of a unitary is equal to the circuit cost, up to polynomial factors and technical caveats [21,22]. Here circuit cost is defined as the length of the

shortest path between two points in curved Riemann geometry. To be specific, for traceless Hermitian operators h_1, h_2, \dots, h_m supported on two qudits and normalized as $\|h\|_{\infty} = 1$, the circuit cost of a unitary $U \in \mathcal{M}_n$ is defined by $\mathcal{C}(U) := \inf_{r_j: [0,1] \rightarrow \mathbb{R}} \int_0^1 \sum_{j=1}^m |r_j(s)| ds$ where r_j satisfies $H(s) = \sum_{j=1}^m r_j(s)h_j$ and $U = P \exp[-i \int_0^1 H(s) ds]$ with the path-ordering operator P . Another important complexity measure is experiment cost, which shows the quantum limit on converting quantum resources, including energy and time, to computational resources [23]. For a circuit with a sequence of gates $U = \Pi_k U_k$, $U_k = \exp(-iH_k T_k)$, where T_k is the runtime of the k th gate, and the time-independent Hamiltonian acting on m_k qubits with spectral decomposition $H_k = \sum_{i=1}^{2^{m_k}} h_i |h_i\rangle\langle h_i|$, $h_i \geq h_{i-1}$. Using the seminorm $E_k := (h_{2^{m_k}} - h_1)/2$, the experiment cost of implementing the gate U_k from the perspective of physical resources is defined as $\mathcal{R}_U := \sum_k \mathcal{R}_{U_k} = \sum_k m_k E_k T_k$. Recently, the lower bounds for circuit cost and experiment cost have been obtained in terms of the quantum Wasserstein complexity measure [11]. The fact will be used in Sec. V.

III. DEFINITION, PROPERTIES, AND CALCULATION OF QUANTUM W_1 DISTANCE BETWEEN UNITARY OPERATIONS

In this section we propose the quantum W_1 distance between unitary operations U, V by the quantum W_1 norm. In Sec. III A we present some properties of the distance. In Sec. III B we show some analytical calculations, including the distance between the identity and some widely used unitary operations.

We show the definition of $\mathcal{D}(U, V)$. It is given by taking the maximization over all input states in terms of the W_1 distance in Definition 2.

Definition 6. Given two unitary operations $U, V \in \mathcal{U}_n$, their quantum Wasserstein distance $\mathcal{D}(U, V) : \mathcal{U}_n \times \mathcal{U}_n \rightarrow \mathbb{R}$ is the maximal quantum W_1 distance between the states they have performed on,

$$\mathcal{D}(U, V) = \max_{\rho \in \mathcal{S}_n} \|U\rho U^\dagger - V\rho V^\dagger\|_{W_1}. \quad (3)$$

By the convexity of the quantum W_1 norm, we need only take the maximization over all pure states,

$$\mathcal{D}(U, V) = \max_{|\psi\rangle\langle\psi| \in \mathcal{S}_n} \|U|\psi\rangle\langle\psi|U^\dagger - V|\psi\rangle\langle\psi|V^\dagger\|_{W_1}, \quad (4)$$

where the maximum is over all normalized pure states $|\psi\rangle \in \mathcal{S}_n$.

A. Properties of $\mathcal{D}(U, V)$

For the convenience of deriving the applications of the quantum W_1 distance between operations, we present some basic properties of $\mathcal{D}(U, V)$.

Proposition 7. The quantum Wasserstein distance $\mathcal{D}(U, V)$ between unitary operations U and V satisfies the following properties:

- (1) Faithfulness: $\mathcal{D}(U, V) = 0$ if and only if $U = V$;
- (2) Symmetry: $\mathcal{D}(U, V) = \mathcal{D}(V, U)$;
- (3) Triangle inequality: $\mathcal{D}(U, V) \leq \mathcal{D}(U, M) + \mathcal{D}(M, V)$, for $U, V, M \in \mathcal{U}_n$;

- (4) Right unitary invariance: $\mathcal{D}(UM, VM) = \mathcal{D}(U, V)$, for $U, V, M \in \mathcal{U}_n$;
 (5) $\mathcal{D}(NU, NV) = \mathcal{D}(U, V)$, for $U, V \in \mathcal{U}_n$ and $N \in \mathcal{U}_1^{\otimes n}$;
 (6) Bounds: $0 \leq \mathcal{D}(U, V) \leq n$, for $U, V \in \mathcal{U}_n$;
 (7) Conjugate transpose invariance with identity: $\mathcal{D}(I, U) = \mathcal{D}(I, U^\dagger)$;
 (8) $\mathcal{D}(U_2 U_1, V_2 V_1) \leq \mathcal{D}(U_1^\dagger, U_2) + \mathcal{D}(V_1^\dagger, V_2)$;
 (9) Superadditivity under tensorization: $\mathcal{D}(U_1 \otimes U_2, V_1 \otimes V_2) \geq \mathcal{D}(U_1, V_1) + \mathcal{D}(U_2, V_2)$;
 (10) $\mathcal{D}(U_1 \otimes U_2, V_1 \otimes V_2) \leq \mathcal{D}(U_1 \otimes I, V_1 \otimes I) + \mathcal{D}(I \otimes U_2, I \otimes V_2)$.

The proof of Proposition 7 is shown in Appendix A.

In addition to basic properties, the W_1 distance shows two physical properties from the perspective of circuit complexity and local distinguishability of unitary operations. The quantum W_1 distance shows the explanation for quantum circuit complexity [11], i.e., the distance $\mathcal{D}(U, V)$ shows the lower bound for the minimum number of gates (smallest circuit) that is required to transform operations U and V to each other. This property of $\mathcal{D}(U, V)$ will be explained in Sec. V. On the other hand, the quantum W_1 distance between operations characterizes the local distinguishability of operations in multiqubit scenarios. Other unitarily invariant distances cannot show such a property [7]. This viewpoint is supported by calculation results, and some comparative examples will be shown in the last paragraph of Sec. III B.

B. The analytical calculation of $\mathcal{D}(U, V)$

By the definition of quantum W_1 distance between states, calculating the distance analytically is a challenge [7]. The analytical calculation of the W_1 distance between operations is more challenging than that of two states. We show the analytical results of the quantum W_1 distance between single-qudit operations, some two-qubit operations, and the multiqubit operations. Their proof is given in Appendix B.

Although the quantum W_1 distance shows benefits mainly in multiqubit scenarios, it is worth mentioning the quantum W_1 distance between arbitrary single-qudit operations.

Proposition 8. The quantum W_1 distance between single-qudit operations U, V shows

(1) For U, V in two-dimensional Hilbert space, it holds that $\mathcal{D}(U, V) = \sqrt{\frac{1}{2}(1 - \cos \alpha)}$, where $\{1, e^{i\alpha}\}$ are the eigenvalues of VU^\dagger .

(2) For U, V in d -dimensional Hilbert space with $d > 2$, it holds that $\mathcal{D}(U, V) = \sqrt{1 - \min_{\sum_{j=0}^{d-1} p_j = 1} |\sum_j p_j e^{i\alpha_j}|^2}$, where $p_j \geq 0$ and $\{e^{i\alpha_0}, e^{i\alpha_1}, \dots, e^{i\alpha_{d-1}}\}$ are the eigenvalues of VU^\dagger .

Remark 9. According to the results presented in Ref. [24],

$$\min_{\sum_j p_j = 1} \left| \sum_j p_j e^{i\alpha_j} \right| = \begin{cases} \cos \frac{\Theta(U^\dagger V)}{2} & 0 \leq \Theta(U^\dagger V) < \pi, \\ 0 & \Theta(U^\dagger V) \geq \pi, \end{cases} \quad (5)$$

where $\Theta(U^\dagger V)$ denotes the length of the smallest arc containing all the eigenvalues of unitary operation $U^\dagger V$ on the unit circle.

Next we consider the W_1 distance between two-qubit unitary operations U and V . Using Property 4 it holds that $\mathcal{D}(U, V) = \mathcal{D}(I, VU^\dagger)$, which implies that it is of great

importance to consider $\mathcal{D}(I, M)$, where M is a unitary operation. First, the W_1 distance between I and the controlled-phase gate is considered. Let $U_\theta^{(k)}$ be a two-qubit diagonal operation whose k th diagonal entry is $e^{i\theta}$ and other diagonal entries are 1, for $k = 1, 2, 3, 4$. We have the following fact.

Proposition 10. The quantum W_1 distance between I and the gate $U_\theta^{(k)}$ is equal to $\sqrt{2} \sin \frac{\theta}{2}$, i.e., $\mathcal{D}(I, U_\theta^{(k)}) = \sqrt{2} \sin \frac{\theta}{2}$.

The CNOT and controlled-Z gate are widely used controlled gate in computation. One can see that the controlled-Z gate $U_{CZ} = U_\pi^{(4)}$ in Proposition 10. On the other hand, it holds that $U_{CZ} = (I \otimes H)U_{CN}(I \otimes H)$. We have the following fact based on the two facts and the single-qubit unitary invariance of $\|\cdot\|_{W_1}$.

Corollary 11. The quantum W_1 distance between I and controlled-Z and CNOT gate is equal to $\sqrt{2}$, i.e., $\mathcal{D}(I, CZ) = \mathcal{D}(I, \text{CNOT}) = \sqrt{2}$.

The operations M described by four-order permutation matrices are also widely used in quantum computation, for example, SWAP gate. We consider $\mathcal{D}(I, M)$ and the following result is obtained.

Proposition 12. Any two-qubit unitary gates switching $|0, 1\rangle|0, 1\rangle$ to $|1, 0\rangle|1, 0\rangle$, or equivalently $|1, 0\rangle|1, 0\rangle$ to $|0, 1\rangle|0, 1\rangle$, have the same quantum Wasserstein distance with the identity I , i.e., $\mathcal{D}(I, U_k) = 2$, where

$$U_1 = \begin{bmatrix} 0 & * & * & * \\ 0 & * & * & * \\ 0 & * & * & * \\ 1 & 0 & 0 & 0 \end{bmatrix}, \quad U_2 = \begin{bmatrix} * & 0 & * & * \\ * & 0 & * & * \\ 0 & 1 & 0 & 0 \\ * & 0 & * & * \end{bmatrix}, \quad (6)$$

$$U_3 = \begin{bmatrix} * & * & 0 & * \\ 0 & 0 & 1 & 0 \\ * & * & 0 & * \\ * & * & 0 & * \end{bmatrix}, \quad U_4 = \begin{bmatrix} 0 & 0 & 0 & 1 \\ * & * & * & 0 \\ * & * & * & 0 \\ * & * & * & 0 \end{bmatrix}. \quad (7)$$

For any unitary operations U, V , the unitary operations M satisfying $M = (U \otimes V)U_k(U \otimes V)^\dagger$ show the same distance with identity I , i.e., $\mathcal{D}(I, M) = 2$.

Obviously, the SWAP gate is included in Proposition 12.

Corollary 13. The quantum W_1 distance between I and the SWAP gate is equal to 2, i.e., $\mathcal{D}(I, \text{swap}) = 2$.

Finally, we show a fact considering the W_1 distance between I and a multiqubit operation. It shows the local discrimination of quantum operations, which is a characteristic property of the quantum W_1 norm between operations.

Proposition 14. For an n -qudit operation consisted of tensor product of k Pauli gate X and $n - k$ identity I , the quantum W_1 distance between it and identity I is equal to k , i.e., $\mathcal{D}(I^{\otimes n}, I^{\otimes(n-k)} \otimes X^{\otimes k}) = k$ for $k = 1, 2, \dots, n$, up to permutations of the qudits.

Using the calculation results above, we illustrate that $\mathcal{D}(U, V)$ characterizes the local distinguishability of operations. Further we compare $\mathcal{D}(U, V)$ with some unitarily invariant distance in the following examples.

(1) Local operations: From Proposition 14, we have $\mathcal{D}(I^{\otimes 2}, I \otimes X) = 1$ and $\mathcal{D}(I^{\otimes 2}, X^{\otimes 2}) = 2$. Adding the Pauli gate X on the second qudit increases the W_1 distance between identity and the total operation. Such a local distinction cannot be detected by unitarily invariant distance, as unitary invariance makes the distance between any couple of quantum states with orthogonal supports

maximal. In particular, one has $\max_{\rho} \frac{1}{2} \|\rho - (I \otimes X)\rho(I \otimes X)\|_1 = \max_{\rho} \frac{1}{2} \|\rho - (X \otimes X)\rho(X \otimes X)\|_1 = 1$, and $\min_{\rho} F(\rho, (I \otimes X)\rho(I \otimes X)) = \min_{\rho} F(\rho, (X \otimes X)\rho(X \otimes X)) = 0$.

(2) Nonlocal operations: We consider two nonlocal gates CNOT and $Q = (I \otimes \sigma_x)U_{CN}(\sigma_x \otimes I)$ differing locally from CNOT gate. From Propositions 11 and 12, we have $\mathcal{D}(I, \text{cnot}) = \sqrt{2}$ and $\mathcal{D}(I, Q) = 2$. Their local distinguishability cannot be described by trace distance fidelity, as we have $\max_{\rho} \frac{1}{2} \|\rho - U_{CN}\rho U_{CN}\|_1 = \max_{\rho} \frac{1}{2} \|\rho - Q\rho Q^{\dagger}\|_1 = 1$ and $\min_{\rho} F(\rho, U_{CN}\rho U_{CN}) = \min_{\rho} F(\rho, Q\rho Q^{\dagger}) = 0$.

IV. ESTIMATION OF THE CLOSENESS BETWEEN OPERATIONS IN QUANTUM CIRCUIT

In this section we show that the W_1 distance between unitary operations is vital in estimating the closeness between operations in quantum circuits. A small $\mathcal{D}(U, V)$ implies that any measurement performed on the states $U|\psi\rangle$ shows approximately the same measurement statistics as that of $V|\psi\rangle$, so U and V plays almost the same role in quantum circuits. The noisy operation simulates the ideal one well when they become close under the W_1 distance.

In a noisy scenario, we denote the ideally and actually implemented unitary operations by U and V , respectively. The W_1 distance between them characterizes how close their measurement outcomes will be in terms of POVM. It is realized by deriving an upper bound of the difference in probability between measurement outcomes.

Proposition 15. Given two operations U and V performed on the same initial state $|\psi\rangle$. Let $M_m \geq 0$ be an element in a POVM performed on $U|\psi\rangle$ and $V|\psi\rangle$, with $P_U^{(m)}$ and $P_V^{(m)}$ being the probability of obtaining the outcome m in the measurements, respectively. The difference between $P_U^{(m)}$ and $P_V^{(m)}$ is upper bounded by the quantum W_1 distance between U and V as

$$|P_U^{(m)} - P_V^{(m)}| \leq 2\lambda_0(M_m)\mathcal{D}(U, V),$$

where $\lambda_0(M_m) \in (0, 1]$ is the maximal eigenvalue of M_m .

Proof. Since $P_U^{(m)}$ and $P_V^{(m)}$ is the probability of obtaining the measurement outcome m , we have

$$|P_U^{(m)} - P_V^{(m)}| = |\langle \psi | U^{\dagger} M_m U | \psi \rangle - \langle \psi | V^{\dagger} M_m V | \psi \rangle|.$$

The POVM operation M_m is positive with the unique positive square root, denoted by N_m , i.e., $M_m = N_m N_m^{\dagger}$ and $\sum_m M_m = I$. Hence,

$$\begin{aligned} & |P_U^{(m)} - P_V^{(m)}| \\ &= |\text{Tr}[N_m^{\dagger}(U|\psi\rangle\langle\psi|U^{\dagger}) - V|\psi\rangle\langle\psi|V^{\dagger})N_m]| \\ &= \left| \sum_k \lambda_k(N_m^{\dagger}(U|\psi\rangle\langle\psi|U^{\dagger}) - V|\psi\rangle\langle\psi|V^{\dagger})N_m \right| \\ &\leq \sum_k |\lambda_k(N_m^{\dagger}(U|\psi\rangle\langle\psi|U^{\dagger}) - V|\psi\rangle\langle\psi|V^{\dagger})N_m| \\ &= \|N_m^{\dagger}(U|\psi\rangle\langle\psi|U^{\dagger}) - V|\psi\rangle\langle\psi|V^{\dagger})N_m\|_1, \end{aligned}$$

where $\lambda_k(X)$ denotes the k th eigenvalue of the operator X , and $\lambda_0 \geq \lambda_1 \geq \dots \geq \lambda_k \dots$. The last equality comes

from $\|X\|_1 = \sum_k |\lambda_k(X)|$ for normal operators. Using the fact that $\|ABC\|_1 \leq \|A\|_{\infty}\|B\|_1\|C\|_{\infty}$ and $\|\rho - \sigma\|_1 \leq 2\|\rho - \sigma\|_{w_1}$, we have

$$\begin{aligned} & \|N_m^{\dagger}(U|\psi\rangle\langle\psi|U^{\dagger}) - V|\psi\rangle\langle\psi|V^{\dagger})N_m\|_1 \\ &\leq \|N_m^{\dagger}\|_{\infty}\|U|\psi\rangle\langle\psi|U^{\dagger}) - V|\psi\rangle\langle\psi|V^{\dagger})\|_1\|N_m\|_{\infty} \\ &\leq 2\|N_m^{\dagger}\|_{\infty}\|N_m\|_{\infty}\|U|\psi\rangle\langle\psi|U^{\dagger}) - V|\psi\rangle\langle\psi|V^{\dagger})\|_{w_1} \\ &\leq 2\|N_m^{\dagger}\|_{\infty}\|N_m\|_{\infty}\mathcal{D}(U, V) \\ &= 2\lambda_0(M_m)\mathcal{D}(U, V), \end{aligned}$$

where the last equality follows from $\|N_m\|_{\infty} = s_0(N_m) = \sqrt{\lambda_0(M_m)}$. Here $s_0(X)$ denotes the maximal singular value of operator X . ■

Proposition 15 implies that if a kind of noise takes the ideal operation to another one and they are close under the W_1 distance, then the noise has little effect on the ideal operation. From the perspective of unitary operation discrimination, a small $\mathcal{D}(U, V)$ also implies that U and V cannot be perfectly distinguished.

In a quantum circuit, the realization of target operations always includes a sequence of unitary gates. So it is important to obtain the distance between two sequences of gates. In analogy to quantifying the distance between an entangled state and a product state, one may be interested in the distance between nonlocal quantum gates and tensor product gates.

Proposition 16. Two sequences of multiqubit unitary gates $U_t U_{t-1} \dots U_1$ and $V_t V_{t-1} \dots V_1$ acting on \mathcal{S}_n , where where V_j can be decomposed as the tensor product of single-qubit gates, for $j = 1, 2, \dots, t$. The quantum W_1 distance between them adds at most linearly with respect to the distance of each couple of gates,

$$\mathcal{D}(U_t U_{t-1} \dots U_1, V_t V_{t-1} \dots V_1) \leq \sum_{k=1}^t \mathcal{D}(U_k, V_k).$$

Proof. We prove it by induction. First we show the case for $t = 2$,

$$\begin{aligned} \mathcal{D}(U_2 U_1, V_2 V_1) &\leq \max_{\rho} \|U_2(U_1 \rho U_1^{\dagger})U_2^{\dagger} - V_2 U_1 \rho U_1^{\dagger} V_2^{\dagger}\|_{w_1} \\ &\quad + \max_{\rho} \|V_2 U_1 \rho U_1^{\dagger} V_2^{\dagger} - V_2 V_1 \rho V_1^{\dagger} V_2^{\dagger}\|_{w_1} \\ &= \mathcal{D}(U_2, V_2) + \mathcal{D}(V_2 U_1, V_2 V_1). \end{aligned}$$

Using property, one has $\mathcal{D}(V_2 U_1, V_2 V_1) = \mathcal{D}(U_1, V_1)$. So we have $\mathcal{D}(U_2 U_1, V_2 V_1) \leq \mathcal{D}(U_2, V_2) + \mathcal{D}(U_1, V_1)$. Suppose the case for $t - 1$ holds, i.e., $\mathcal{D}(U_{t-1} \dots U_1, V_{t-1} \dots V_1) \leq \sum_{k=1}^{t-1} \mathcal{D}(U_k, V_k)$. Then we have

$$\begin{aligned} & \mathcal{D}(U_t U_{t-1} \dots U_1, V_t V_{t-1} \dots V_1) \\ &\leq \mathcal{D}(U_t, V_t) + \mathcal{D}(U_{t-1} U_{t-2} \dots U_1, V_{t-1} V_{t-2} \dots V_1) \\ &\leq \sum_{k=1}^t \mathcal{D}(U_k, V_k), \end{aligned}$$

which is the desired result. ■

Propositions 15 and 16 can be applied to estimate the measurement outcome of the circuits containing different sequences of gates U_1, U_2, \dots, U_t and V_1, V_2, \dots, V_t . In practice, we set a tolerance $\alpha > 0$ of the probability that two

circuits show the same measurement outcome. We can estimate how close the effects of these gates are in the circuits, i.e., whether the probability of different measurement outcomes is within the tolerance, only by the distance $\mathcal{D}(U_k, V_k)$. Specifically, to make the probability of different measurement outcomes be within the tolerance α , it suffices that

$$|P_{U_1 \dots U_t}^{(m)} - P_{V_1 \dots V_t}^{(m)}| \leq 2\lambda_0(M_m) \sum_{k=1}^t \mathcal{D}(U_k, V_k) \leq \alpha, \quad (8)$$

where $P_{U_1 \dots U_t}^{(m)}$, $P_{V_1 \dots V_t}^{(m)}$ and $\lambda_0(M_m)$ have been defined in Proposition 15. The inequality (8) holds when

$$\mathcal{D}(U_k, V_k) \leq \frac{\alpha}{2t \max_m \{\lambda_0(M_m)\}} \quad (9)$$

for $k = 1, 2, \dots, t$. An example of the above process is given.

Example 17. A sequence of ideal qubit gates U_1, U_2, \dots, U_5 in the quantum circuit is subject to the unitary noise process $\mathcal{E}_\theta = \text{diag}\{e^{i\theta}, e^{-i\theta}\}$, where $\theta \in [0, \pi]$ is the parameter related to noise. The ideal gates are transformed into a sequence of noisy gates V_1, V_2, \dots, V_5 , where

$$U_k = \begin{bmatrix} e^{\alpha_k i} & 0 \\ 0 & e^{\beta_k i} \end{bmatrix}, \quad V_k = U_k \mathcal{E}_\theta = \begin{bmatrix} e^{(\alpha_k + \theta)i} & 0 \\ 0 & e^{(\beta_k - \theta)i} \end{bmatrix}$$

for $k = 1, 2, \dots, 5$. Using Proposition 8, one has $\mathcal{D}(U_k, V_k) = |\sin \theta|$. Suppose the following POVM $\{M_m : m = 1, 2, \dots, 8\}$ is carried out in the circuit,

$$\begin{aligned} M_1 &= \begin{bmatrix} \frac{1}{8} & -\frac{1}{8}i \\ \frac{1}{8}i & \frac{1}{8} \end{bmatrix}, & M_2 &= \begin{bmatrix} \frac{1}{8} & \frac{1}{8}i \\ -\frac{1}{8}i & \frac{1}{8} \end{bmatrix}, & M_3 &= \begin{bmatrix} \frac{1}{8} & \\ \frac{1}{8} & \frac{1}{8} \end{bmatrix}, \\ M_4 &= \begin{bmatrix} \frac{1}{8} & -\frac{1}{8} \\ -\frac{1}{8} & \frac{1}{8} \end{bmatrix}, & M_5 &= \begin{bmatrix} \frac{1}{8} & \frac{1}{8}e^{\pi i/4} \\ \frac{1}{8}e^{-\pi i/4} & \frac{1}{8} \end{bmatrix}, \\ M_6 &= \begin{bmatrix} \frac{1}{8} & \frac{1}{8}e^{5\pi i/4} \\ \frac{1}{8}e^{-5\pi i/4} & \frac{1}{8} \end{bmatrix}, & M_7 &= \begin{bmatrix} \frac{1}{4} & 0 \\ 0 & 0 \end{bmatrix}, \\ M_8 &= \begin{bmatrix} 0 & 0 \\ 0 & \frac{1}{4} \end{bmatrix}. \end{aligned}$$

We set the probability error tolerance $\alpha = 30\%$. To make the probability of different measurement outcomes be within the tolerance α for any initial state $|\psi\rangle$, i.e., $|P_{U_1 \dots U_t}^{(m)} - P_{V_1 \dots V_t}^{(m)}| \leq \alpha$, it suffices that

$$\mathcal{D}(U_k, V_k) = |\sin \theta| \leq \frac{\alpha}{10 \max_m \{\lambda_0(M_m)\}} = 0.12.$$

It implies that each noisy gate V_k simulates the ideal gate U_k within the tolerance effectively if the parameter of local noise $\theta \in [0, \arcsin(0.12)]$ or $\theta \in [\pi - \arcsin(0.12), \pi]$.

V. THE W_1 GATE ERROR RATE UNDER NOISE

In this section we introduce a measure quantifying the effect of noise on quantum gates, named the W_1 gate error rate. We establish the relation between the W_1 error rate and two real cost measures of recovery operation, including circuit and experiment costs. Further we show two examples considering the implementation under depolarizing and unitary noise for arbitrary single-qubit gate and CNOT gate, respectively.

Following the idea of proposing the gate error rate in Sec. II C, we define the error rate of a quantum gate as follows.

Definition 18. The W_1 error rate of the implementation of n -qubit unitary gate U is given by

$$e(U, \mathcal{V}) := \frac{1}{n} \max_{\rho} \|U \rho U^\dagger - \mathcal{V}(\rho)\|_{W_1}, \quad (10)$$

where $\rho \in \mathcal{S}_n$ and \mathcal{V} is a channel that describes the noisy implementation of U .

For two states $\rho, \sigma \in \mathcal{S}_n$, we have $\|\rho - \sigma\|_{W_1} \in [0, n]$ by Definition 2 and hence the error rate $e(U, \mathcal{V}) \in [0, 1]$. Compared with the error rate induced by Schatten 1-norm and diamond norm in Sec. II C, the following relation can be obtained, $e(U, \mathcal{V}) \leq e_1(U, \mathcal{V}) \leq e_\diamond(U, \mathcal{V})$, where the first inequality comes from Lemma 2, and the second one follows directly from their definitions.

We show the benefits of W_1 error rate from the perspective of QEC and circuit complexity. These results imply that $e(U, \mathcal{V})$ is a figure of merit quantifying the effect of noise on a specific unitary operation, from the perspective of experiment cost to recover the influence of noise. We consider the noise process described by the channel $\mathcal{V} = \mathcal{G} \circ \mathcal{E}$, where $\mathcal{G}(\cdot) = U(\cdot)U^\dagger$ with $U \in \mathcal{U}_n$ denotes the ideal implementation of gate, and $\mathcal{E}(\cdot) = \sum_{k=1}^N p_k V_k(\cdot)V_k^\dagger$ denotes a general noise process described by the mixed unitary channel [25], with a probability vector (p_1, p_2, \dots, p_N) and $V_1, V_2, \dots, V_N \in \mathcal{U}_n$. The error rate of gate U and its upper bound is given by

$$\begin{aligned} e(U, \mathcal{V}) &= \frac{1}{n} \max_{\rho} \left\| U \rho U^\dagger - U \left(\sum_{k=1}^N p_k V_k \rho V_k^\dagger \right) U^\dagger \right\|_{W_1} \quad (11) \\ &\leq \frac{1}{n} \sum_k p_k \mathcal{D}(I, UV_k U^\dagger), \quad (12) \end{aligned}$$

where the inequality follows from the convexity of $\|\cdot\|_{W_1}$. From (12), analyzing the upper bound of $e(U, \mathcal{V})$ suffices to consider the noise process described by each unitary error V_k . We set the noise channel $\mathcal{E}_{V_k}(\cdot) = V_k(\cdot)V_k^\dagger$ and noisy implementation of gate $\mathcal{V}_{V_k} = \mathcal{G} \circ \mathcal{E}_{V_k} = UV_k(\cdot)V_k^\dagger U^\dagger$. Using Properties 4 and 7, one has $e(U, \mathcal{V}_{V_k}) = \frac{1}{n} \mathcal{D}(I, P^{V_k})$, where $P^{V_k} = UV_k^\dagger U^\dagger$ is the recovery operation of ideal gate U under noise \mathcal{V}_{V_k} , as performing P^{V_k} on the noisy gate UV_k can correct the influence of noise, i.e., $P^{V_k}(UV_k) = U$.

As we all know, quantum error correction (QEC) includes the error detection and recovery step [19]. Recovery operations in the second step are to eliminate the noise on specific qubits, such as $P^{V_k} = UV_k^\dagger U^\dagger$ above. Compared with the existing error rate introduced in Sec. II C, the W_1 error rate has an explanation in terms of the physical resource cost for recovery operations. Since $\mathcal{D}(U, V)$ characterizes local distinguishability, the local alteration for operations will proportionally change the total distance, while no unitarily invariant distance has this property [7,26]. For example, an ideal gate $U_{id} = \sigma_x \otimes \sigma_x$ is performed on $\rho = |00\rangle\langle 00|$. Two noisy implementation of U_{id} shows $U_{ac}^{(1)} = I^{\otimes 2}$ and $U_{ac}^{(2)} = I \otimes \sigma_x$, which generates the resulting state $\rho_{ac}^{(k)}$. The distance $d(\cdot, \cdot)$ induced by unitary-invariant norms or fidelity shows $d(U_{id}, U_{ac}^{(1)}) = d(U_{id}, U_{ac}^{(2)})$, while $\mathcal{D}(U_{id}, U_{ac}^{(1)}) = 2$, and $\mathcal{D}(U_{id}, U_{ac}^{(2)}) = 1$. In the recovery step of QEC, two gates $\sigma_x \otimes \sigma_x$ are required for $\rho_{ac}^{(1)}$, and only one σ_x is required for

$\rho_{ac}^{(2)}$. So $U_{ac}^{(1)}$ is farther away from U_{id} than $U_{ac}^{(2)}$ in terms of experiment resource, which is characterized by W_1 distance and error rate.

Next we introduce a fact considering the quantum W_1 distance and circuit complexity. The main results of Ref. [11] show the lower bounds for circuit cost $\mathcal{C}(U)$ and experiment cost $\mathcal{R}(U)$ for a unitary operation U . We rephrase them by quantum W_1 distance as follows:

$$\mathcal{C}(U) \geq 4\sqrt{2}\mathcal{D}(I, U), \quad (13)$$

$$\mathcal{R}(U) \geq \frac{1}{2}\mathcal{D}(I, U). \quad (14)$$

It shows that $\mathcal{D}(I, U)$ provides the lower bounds for the minimum number of basic gates and resources to realize the operation U . Hence, $\mathcal{D}(U, V)$ provides the lower bounds to transform the unitary operation U to V .

We establish the relation between W_1 error rate and circuit complexity based on the above fact. We choose U as the recovery operation P^{V_k} , and thus $\mathcal{C}(P^{V_k})$ and $\mathcal{R}(P^{V_k})$ are the practical cost to recover the influence of noise \mathcal{V}_k on the ideal operation U . We consider all possible unitary errors in the mixed unitary channel $\mathcal{E}(\cdot)$. Combined with (12)–(14), it can be obtained that

$$e(U, \mathcal{V}) \leq \sum_k \frac{p_k}{4\sqrt{2n}} \mathcal{C}(P^{V_k}), \quad (15)$$

$$e(U, \mathcal{V}) \leq \sum_k \frac{2p_k}{n} \mathcal{R}(P^{V_k}). \quad (16)$$

It implies that the W_1 error rate $e(U, \mathcal{V})$ provides a lower bound of circuit and experiment cost for recovery operation of U under arbitrary noise \mathcal{V} . That is to say, $e(U, \mathcal{V})$ is related to the minimal quantum resources required to eliminate the influence of noise $\mathcal{E}(\cdot)$ during the realization of operation U . Thus, the W_1 error rate is a figure of merit concerning the noisy gate and the experimental requirement to eliminate noise acting on the gate. So it quantifies the effect of noise from the perspective of physical resources.

Finally, we show two examples of noisy implementation of CNOT gates.

Example 19. We consider the depolarizing noise and unitary noise acting on a single qubit. The noise process is given by the channels respectively, $\mathcal{E}_p^{\text{dep},1}(\rho) := (1-p)\rho + p\frac{\mathbb{I}_2}{2}$, $\mathcal{E}_\theta^{\text{uni}}(\rho) := E_\theta\rho E_\theta^\dagger$, where $p \in [0, 1]$ and E_θ is a unitary operator with eigenvalues $e^{\pm\theta i}$ for $\theta \in [0, \pi]$. The error rate of a single-qubit gate U shows

$$e(U, \mathcal{V}_{\text{dep},1}) = \frac{p}{2}, \quad e(U, \mathcal{V}_{\text{uni}}) = \sqrt{1 - \cos 2\theta}. \quad (17)$$

The average gate fidelities for depolarizing noise and unitary noise are, respectively [18],

$$\varphi_p^{\text{dep}} = 1 - \frac{p}{2}, \quad \varphi_\theta^{\text{uni}} = \frac{1}{3} + \frac{2}{3}\cos^2\theta. \quad (18)$$

The error rate induced by the diamond norm is [18]

$$e_\diamond(U, \mathcal{V}_{\text{dep},1}) = \frac{3}{4}p, \quad e_\diamond(U, \mathcal{V}_{\text{uni}}) = \sin\theta. \quad (19)$$

Generally, the advantage of quantum W_1 norm appears for the multiqubit case. We show the example for the W_1 error rate

of noisy implementation of the CNOT gate in the presence of two typical kinds of noise.

Example 20. We consider the noisy implementation of the CNOT gate under unitary and depolarizing noise as follows.

First, we consider the W_1 error rate of noisy implementation of the CNOT gate under unitary noise channel $\mathcal{E}_{CP}(\rho) = U_{CP}\rho U_{CP}^\dagger$, for $U_{CP} = \text{diag}\{1, 1, 1, e^{i\theta}\}$, $\theta \in [0, 2\pi)$. We denote the actual implementation of the CNOT gate under unitary noise as $\mathcal{V}_{\text{uni},2} = \mathcal{G}_{CN} \circ \mathcal{E}_{CP}$, for $\mathcal{G}_{CN}(\cdot) = U_{CN}(\cdot)U_{CN}$. From Proposition 10, we have

$$e(\text{CNOT}, \mathcal{V}_{\text{uni},2}) = \frac{1}{2}\mathcal{D}(I, U_{CN}U_{CP}U_{CN}) = \frac{1}{\sqrt{2}}\sin\frac{\theta}{2}.$$

Using (15), the lower bounds for circuit cost and experiment cost are, respectively,

$$\mathcal{C}(U_{CN}U_{CP}U_{CN}) = 8\sin\frac{\theta}{2}, \quad (20)$$

$$\mathcal{R}(U_{CN}U_{CP}U_{CN}) = \frac{1}{\sqrt{2}}\sin\frac{\theta}{2}. \quad (21)$$

It is the minimum quantum resource required to eliminate the influence of noise $\mathcal{V}_{\text{uni},2}$ during the implementation of the CNOT gate.

Next, we consider the depolarizing channel acting on \mathcal{S}_2 as $\mathcal{E}_p^{\text{dep},2}(\rho) := (1-p)\rho + p\frac{\mathbb{I}_4}{4} = (1-p)\rho + \frac{p}{16}\sum_{s,t=0}^3(X^s Z^t)(\rho)(X^s Z^t)^\dagger$, where $X = \sum_{q=0}^3|q \oplus 1\rangle\langle q|$ and $Z = \sum_{q=0}^3 i^q|q\rangle\langle q|$. We denote $\mathcal{V}_{\text{dep},2} = \mathcal{G}_{CN} \circ \mathcal{E}_p^{\text{dep},2}$ as the actual implementation of the CNOT gate. The W_1 error rate of the CNOT gate can be estimated as follows:

$$e(\text{CNOT}, \mathcal{V}_{\text{dep},2}) = \frac{1}{2}\max_\rho \|\mathcal{G}_{CN}(\rho) - \mathcal{G}_{CN} \circ \mathcal{E}_p^{\text{dep},2}(\rho)\|_{W_1} \in \left[\frac{3}{8}p, \frac{3}{4}p\right], \quad (22)$$

where the range is derived from Lemma 2, i.e., $\frac{3}{8}p = \frac{p}{2}\max_\rho \frac{1}{2}\|U_{CN}\rho U_{CN} - \frac{\mathbb{I}_4}{4}\|_1 \leq e(\text{CNOT}, \mathcal{V}_{\text{dep},2}) \leq \frac{p}{2}\max_\rho \|U_{CN}\rho U_{CN} - \frac{\mathbb{I}_4}{4}\|_1 = \frac{3}{4}p$. Using (12), we have

$$e(\text{CNOT}, \mathcal{V}_{\text{dep},2}) \leq \frac{p}{32}\sum_{s,t=0}^3 \mathcal{D}(I, U_{CN}X^s Z^t U_{CN}). \quad (23)$$

From (15), the average lower bounds of circuit cost and experiment cost concerning the depolarizing noise are

$$\frac{p}{16}\sum_{s,t \neq 0} \mathcal{C}(U_{CN}X^s Z^t U_{CN}) \geq 3\sqrt{2}p, \quad (24)$$

$$\frac{p}{16}\sum_{s,t \neq 0} \mathcal{R}(U_{CN}X^s Z^t U_{CN}) \geq \frac{3}{8}p, \quad (25)$$

respectively.

VI. CONCLUSION

In summary, we have introduced the quantum Wasserstein distance between unitary operations, we presented its properties and analytical calculations. The characterizing properties of this distance is that it characterizes local discrimination and shows an explanation for circuit complexity. The closeness between operations can be estimated in quantum circuits with

W_1 distance. The smaller the distance is, the similar their measurement outcomes will be in the circuit. As an application, we introduced the W_1 error rate by the distance. We showed its estimation and established the relation between the W_1 error rate and two practical cost measures of recovery operations, including circuit cost and experiment cost. We showed two examples considering the implementation under depolarizing and unitary noise for an arbitrary single-qubit gate and CNOT gate, respectively. So the W_1 distance quantifies the effect of noise on unitary operations from the perspective of experiment resource.

Many problems arising from this paper can be further explored. The analytical calculation of the W_1 distance between arbitrary operations is left as an open problem. It may be efficiently approximated by sampling method in [13]. As a similarity measure of operations, the W_1 distance may be employed to design the loss functions in quantum operation learning. On the other hand, we have shown that $\mathcal{D}(I, \text{CNOT}) = \sqrt{2}$ and $\mathcal{D}(I, \text{SWAP}) = 2$. In [27], it has been obtained that the entangling power of CNOT and SWAP gates are one and two ebits, respectively. So this distance may be developed to characterize more properties of unitary operations, such as entangling power.

ACKNOWLEDGMENTS

We thank L. Yu for a careful reading of the paper. L.C. was supported by the NNSF of China (Grant No. 11871089) and the Fundamental Research Funds for the Central Universities (Grant No. ZG216S2005). L.-J.Z. was supported by the NNSF of China (Grants No. 12101031 and No. 11947241)

APPENDIX A: PROOF OF PROPERTIES FOR $\mathcal{D}(U, V)$

Here we show the proof of Proposition 7.

Proof. Properties 1 and 2 follow from the faithfulness and symmetry of the quantum W_1 norm, respectively.

Property 3 follows from the triangle inequality of the quantum W_1 norm,

$$\begin{aligned} \|U\rho U^\dagger - V\rho V^\dagger\|_{W_1} &\leq \|V\rho V^\dagger - M\rho M^\dagger\|_{W_1} \\ &\quad + \|M\rho M^\dagger - U\rho U^\dagger\|_{W_1}. \end{aligned}$$

The equality holds when $M = e^{i\theta}U$, $M = e^{i\phi}V$ or $V\rho V^\dagger - M\rho M^\dagger = k(M\rho M^\dagger - U\rho U^\dagger)$, for $k \geq 0$.

Property 4 can be proved as follows:

$$\begin{aligned} \mathcal{D}(UM, VM) &= \max_{|\psi\rangle} \|UM|\psi\rangle\langle\psi|M^\dagger U^\dagger - VM|\psi\rangle\langle\psi|M^\dagger V^\dagger\|_{W_1} \\ &= \max_{|\xi\rangle} \|U|\xi\rangle\langle\xi|U^\dagger - V|\xi\rangle\langle\xi|V^\dagger\|_{W_1}, \end{aligned}$$

where $|\xi\rangle = M|\psi\rangle$ is a pure state. Thus, the maximization takes over all pure states. The last equality is equal to $\mathcal{D}(U, V)$.

Property 5 can be obtained as the quantum W_1 distance is invariant concerning unitary operations acting on a single qubit.

Property 6 is obtained with Property 4 by choosing $M = U^\dagger$,

$$\begin{aligned} \mathcal{D}(U, V) &\leq \max_{|\psi\rangle} \left\{ \frac{n}{2} \|U|\psi\rangle\langle\psi|U^\dagger - V|\psi\rangle\langle\psi|V^\dagger\|_1 \right\} \\ &= \max_{|\psi\rangle} \{n\sqrt{1 - |\langle\psi|U^\dagger V|\psi\rangle|^2}\} = n, \end{aligned}$$

where the inequality comes from the fact in Lemma 2. On the other hand, $\mathcal{D}(U, V) \geq 0$ is obtained directly from the nonnegativity of the quantum W_1 norm. So the desired result is obtained.

Property 7 is proved as follows:

$$\begin{aligned} \mathcal{D}(I, U) &= \max_{|\psi\rangle} \| |\psi\rangle\langle\psi| - U|\psi\rangle\langle\psi|U^\dagger \|_{W_1} \\ &= \max_{|\xi\rangle=U|\psi\rangle} \| |\xi\rangle\langle\xi| - U^\dagger|\xi\rangle\langle\xi|U \|_{W_1} \\ &= \mathcal{D}(I, U^\dagger). \end{aligned}$$

Property 8 is proved with the help of Properties 3 and 4. One can obtain that

$$\begin{aligned} \mathcal{D}(U_2U_1, V_2V_1) &\leq \max_{\rho} \|U_2U_1\rho U_1^\dagger U_2^\dagger - \rho\|_{W_1} + \|\rho - V_2V_1\rho V_1^\dagger V_2^\dagger\|_{W_1} \\ &= \mathcal{D}(I, U_2U_1) + \mathcal{D}(I, V_2V_1) = \mathcal{D}(U_1^\dagger, U_2) + \mathcal{D}(V_1^\dagger, V_2). \end{aligned}$$

The equality holds when $U_2U_1 = e^{i\theta}I$, $V_2V_1 = e^{i\phi}I$, or $U_2U_1\rho U_1^\dagger U_2^\dagger - \rho = k(\rho - V_2V_1\rho V_1^\dagger V_2^\dagger)$, for $k \geq 0$.

Property 9 holds from the tensorization of the quantum W_1 norm in Lemma 3. We have

$$\begin{aligned} \|(U_1 \otimes U_2)\rho(U_1^\dagger \otimes U_2^\dagger) - (V_1 \otimes V_2)\rho(V_1^\dagger \otimes V_2^\dagger)\|_{W_1} \\ \geq \|U_1\rho_1U_1^\dagger - V_1\rho_1V_1^\dagger\|_{W_1} + \|U_2\rho_2U_2^\dagger - V_2\rho_2V_2^\dagger\|_{W_1}, \end{aligned}$$

where ρ_i denotes the corresponding reduced state. Take the maximum of all pure states ρ, ρ_1, ρ_2 on both sides of the above inequality. Then the property can be proved.

Next, we prove property 10 by the triangle inequality of the W_1 norm. Let $\sigma = (U_1 \otimes U_2)\rho(U_1^\dagger \otimes U_2^\dagger)$ and $\xi = (V_1U_1^\dagger \otimes I)\sigma(U_1V_1^\dagger \otimes I)$. One has

$$\begin{aligned} \|(U_1 \otimes U_2)\rho(U_1^\dagger \otimes U_2^\dagger) - (V_1 \otimes V_2)\rho(V_1^\dagger \otimes V_2^\dagger)\|_{W_1} \\ = \|\sigma - (V_1 \otimes V_2)(U_1^\dagger \otimes U_2^\dagger)\sigma(U_1 \otimes U_2) \\ \times (V_1^\dagger \otimes V_2^\dagger)\|_{W_1} \leq \|\sigma - \xi\|_{W_1} \\ + \|\xi - (I \otimes V_2U_2^\dagger)(V_1U_1^\dagger \otimes I)\sigma(U_1V_1^\dagger \otimes I)(I \otimes U_2V_2^\dagger)\|_{W_1} \\ = \|(U_1 \otimes I)\eta(U_1^\dagger \otimes I) - (V_1 \otimes I)\eta(V_1^\dagger \otimes I)\|_{W_1} \\ + \|(I \otimes U_2)\mu(I \otimes U_2^\dagger) - (I \otimes V_2)\mu(I \otimes V_2^\dagger)\|_{W_1} \\ \leq \mathcal{D}(U_1 \otimes I, V_1 \otimes I) + \mathcal{D}(I \otimes U_2, I \otimes V_2), \end{aligned}$$

where $\eta = (U_1^\dagger \otimes I)\sigma(U_1 \otimes I)$, and $\mu = (I \otimes U_2^\dagger)(V_1U_1^\dagger \otimes I)\sigma(U_1V_1^\dagger \otimes I)(I \otimes U_2)$. Hence it holds that $\mathcal{D}(U_1 \otimes U_2, V_1 \otimes V_2) \leq \mathcal{D}(U_1 \otimes I, V_1 \otimes I) + \mathcal{D}(I \otimes U_2, I \otimes V_2)$. ■

APPENDIX B: THE ANALYTICAL CALCULATION OF W_1 DISTANCE

We present the calculation of the distance between I and some well-known gates. In Sec. B 1 we show the proof of

Proposition 10. In Sec. B 2 we show the proof of Propositions 8, 12, and 14.

1. The W_1 distance between I and generalized controlled phase gate

We denote the generalized controlled phase gate as $U_\theta^{(k)}$ which is the diagonal two-qubit operation whose k th diagonal entry is $e^{i\theta}$ and other diagonal entries are 1, for $k = 1, 2, 3, 4$. First we consider $\mathcal{D}(I, U_\theta^{(3)})$, where U_{CP} and U_{CN} are given in Sec. II A. We rephrase Proposition 10 from Sec. III B for convenience.

Proposition 21. The quantum W_1 distance between I and controlled-phase gate $U_\theta^{(3)} = U_{CN}U_{CP}U_{CN} = \text{diag}\{1, 1, e^{i\theta}, 1\}$ is equal to $\sqrt{2} \sin \frac{\theta}{2}$, i.e., $\mathcal{D}(I, U_{CN}U_{CP}U_{CN}) = \sqrt{2} \sin \frac{\theta}{2}$.

Proof. We calculate $\mathcal{D}(I, U_{CN}U_{CP}U_{CN})$ by finding its upper and lower bounds. If the upper bound coincides with the lower bound, we obtain the desired value.

For a pure state $|\psi\rangle = \sum_{m,n=0}^1 a_{m,n}|m, n\rangle$ with $\sum_{m,n} |a_{m,n}|^2 = 1$, we set the state $\rho = |\psi\rangle\langle\psi|$ and $\sigma = (U_{CN}U_{CP}U_{CN})\rho(U_{CN}U_{CP}U_{CN})^\dagger$. From Corollary 4, it can be obtained that

$$\|\rho - \sigma\|_{W_1} \geq \frac{1}{2}\|\rho_1 - \sigma_1\|_1 + \frac{1}{2}\|\rho_2 - \sigma_2\|_1 \quad (\text{B1})$$

$$= 2 \sin \frac{\theta}{2} |a_{1,0}|(|a_{0,0}| + |a_{1,1}|), \quad (\text{B2})$$

where ρ_k and σ_k are the reduced density operator of ρ and σ , respectively. From $\sum_{m,n} |a_{m,n}|^2 = 1$, we have $2|a_{1,0}|(|a_{0,0}| + |a_{1,1}|) \leq \sqrt{2}$, and the equality holds for the input state $\rho_{\text{inf}} = |\psi_1\rangle\langle\psi_1|$ and $\sigma_{\text{inf}} = (U_{CN}U_{CP}U_{CN})\rho_{\text{inf}}(U_{CN}U_{CP}U_{CN})^\dagger$, where $|\psi_1\rangle = \sum_{m,n} a_{m,n}|m, n\rangle$ with the coefficients satisfying $|a_{0,0}| = |a_{1,1}| = \frac{1}{2}$, $|a_{0,1}| = 0$, $|a_{1,0}| = \frac{1}{\sqrt{2}}$. From Definition 6, $\mathcal{D}(I, U_{CN}U_{CP}U_{CN})$ is obtained by taking the maximization over all input states. From (B1), we have obtained that $\|\rho_{\text{inf}} - \sigma_{\text{inf}}\|_{W_1} \geq \sqrt{2} \sin \frac{\theta}{2}$. Hence we have

$$\mathcal{D}(I, U_{CN}U_{CP}U_{CN}) \geq \sqrt{2} \sin \frac{\theta}{2}. \quad (\text{B3})$$

According to Definition 2, one has

$$\begin{aligned} & \mathcal{D}(I, U_{CN}U_{CP}U_{CN}) \\ &= \max_{\rho \in \mathcal{S}_2} \min \left\{ \sum_{i=1}^2 c_i : c_i \geq 0, \text{Tr}_i F^{(i)} = 0, \right. \\ & \left. \rho - (U_{CN}U_{CP}U_{CN})\rho(U_{CN}U_{CP}U_{CN})^\dagger = \sum_{i=1}^2 c_i F^{(i)} \right\}, \end{aligned} \quad (\text{B4})$$

where $F^{(i)} = \rho^{(i)} - \sigma^{(i)} \in \mathcal{M}_2$ satisfying $\text{Tr}_i F^{(i)} = 0$. Since $\mathcal{D}(I, U_{CN}U_{CP}U_{CN})$ is derived by taking the minimization of $c_1 + c_2$ over all $F^{(i)}$'s, the coefficient $c_1 + c_2$ induced by a particular set of $F^{(1)}$ and $F^{(2)}$ is the upper bound of $\mathcal{D}(I, U_{CN}U_{CP}U_{CN})$. We consider the particular case for $F^{(i)} = \rho^{(i)} - M\rho^{(i)}M \in \mathcal{M}_2$ satisfying $\text{Tr}_i F^{(i)} = 0$, for $M = \text{diag}\{1, 1, -1, 1\}$. Here $\rho^{(i)}$ is any two-qubit pure state. We aim to show that the upper bound of (B4) is equal to $\sqrt{2} \sin \frac{\theta}{2}$.

It can be realized by finding a couple of $F^{(1)}$ and $F^{(2)}$ such that $c_1 + c_2 \leq \sqrt{2} \sin \frac{\theta}{2}$ for all pure states ρ .

For any pure state $\rho = \sum_{j,k,s,t} a_{j,k} a_{s,t}^* |j, k\rangle\langle s, t|$, one has

$$\rho - (U_{CN}U_{CP}U_{CN})\rho(U_{CN}U_{CP}U_{CN})^\dagger = [d_{m,n}]_{4 \times 4},$$

where $d_{1,3} = a_{0,0}a_{1,0}^*(1 - e^{-i\theta})$, $d_{2,3} = a_{0,1}a_{1,0}^*(1 - e^{-i\theta})$, $d_{3,1} = a_{0,0}^*a_{1,0}(1 - e^{i\theta})$, $d_{3,2} = a_{0,1}^*a_{1,0}(1 - e^{i\theta})$, $d_{3,4} = a_{1,0}a_{1,1}^*(1 - e^{i\theta})$, $d_{4,3} = a_{1,0}^*a_{1,1}(1 - e^{-i\theta})$ and other entries are equal to zero. Any $F^{(k)} \in \mathcal{M}_2$ satisfying $\text{Tr}_k F^{(k)} = 0$ can be written as

$$F^{(1)} = \begin{bmatrix} 0 & 0 & 2g_{0,0}g_{1,0}^* & 0 \\ 0 & 0 & 2g_{0,1}g_{1,0}^* & 0 \\ 2g_{0,0}^*g_{1,0} & 2g_{0,1}^*g_{1,0} & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad (\text{B5})$$

$$F^{(2)} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 2h_{0,1}h_{1,0}^* & 0 \\ 0 & 2h_{0,1}^*h_{1,0} & 0 & 2h_{1,0}h_{1,1}^* \\ 0 & 0 & 2h_{1,0}^*h_{1,1} & 0 \end{bmatrix}, \quad (\text{B6})$$

where the entries satisfy $\sum_{j,k} |g_{j,k}|^2 = \sum_{j,k} |h_{j,k}|^2 = 1$.

We need to find the g_{jk} , h_{jk} , and c_k , such that

$$\rho - (U_{CN}U_{CP}U_{CN})\rho(U_{CN}U_{CP}U_{CN})^\dagger = c_1 F^{(1)} + c_2 F^{(2)}, \quad (\text{B7})$$

$$c_1 + c_2 = \sqrt{2} \sin \frac{\theta}{2} \quad (\text{B8})$$

holds for any ρ , i.e.,

$$a_{0,1}a_{1,0}^* \frac{(1 - e^{-i\theta})}{2} = c_1 g_{0,1}g_{1,0}^* + c_2 h_{0,1}h_{1,0}^* \quad (\text{B9a})$$

$$a_{0,0}a_{1,0}^* \frac{(1 - e^{-i\theta})}{2} = c_1 g_{0,0}g_{1,0}^*, \quad (\text{B9b})$$

$$a_{1,1}a_{1,0}^* \frac{(1 - e^{-i\theta})}{2} = c_2 h_{1,1}h_{1,0}^*, \quad (\text{B9c})$$

$$\sum_{j,k} |g_{j,k}|^2 = \sum_{j,k} |h_{j,k}|^2 = \sum_{j,k} |a_{j,k}|^2 = 1, \quad (\text{B9d})$$

$$c_1 + c_2 = \sqrt{2} \sin \frac{\theta}{2}, \quad c_k \geq 0, \quad \text{for } k = 1, 2 \quad (\text{B9e})$$

holds for any $a_{j,k}$. We consider two margin cases for the coefficients $a_{j,k}$, which will be used later.

(1) From (B5), one can obtain that $\|\rho - (U_{CN}U_{CP}U_{CN})\rho(U_{CN}U_{CP}U_{CN})^\dagger\|_{W_1} = 0$ when $a_{1,0} = 0$ or $a_{1,0} = 1$. Using (B4), it can be obtained that $c_1 + c_2 \leq \sqrt{2} \sin \frac{\theta}{2}$.

(2) If $a_{0,0} = a_{1,1} = 0$, we can choose $c_1 = c_2 = \frac{1}{\sqrt{2}} \sin \frac{\theta}{2}$, $g_{0,0} = h_{1,1} = 0$, $g_{1,0}^* = h_{0,1} = \frac{1}{\sqrt{2}} a_{0,1}$, $g_{0,1}^* = h_{1,0} = a_{1,0}$, $\exp\{i(\frac{\theta}{2} - \frac{\pi}{2})\}$ and $g_{1,1} = h_{0,0} = \sqrt{1 - \frac{1}{2}|a_{0,1}|^2 - |a_{1,0}|^2}$, such that (B9a)-(B9e) hold.

Next, we only consider the case for

$$a_{1,0} \neq 0, 1, \quad a_{0,0} \neq 0 \quad \text{or} \quad a_{1,1} \neq 0. \quad (\text{B10})$$

Using the normalization condition in (B9d), we can parametrize the coefficients $g_{j,k}$ and $h_{j,k}$ by $\alpha_j, \beta_j \in [0, \frac{\pi}{2}]$,

$$g_{0,0} = \cos \alpha_0 \sin \alpha_1 e^{i\alpha_2}, \quad g_{0,1} = \cos \alpha_0 \cos \alpha_1, \quad (\text{B11})$$

$$g_{1,0} = \sin \alpha_0 \sin \alpha_3 e^{i\alpha_6}, \quad g_{1,1} = \sin \alpha_0 \cos \alpha_3 e^{i\alpha_4}, \quad (\text{B12})$$

$$h_{0,0} = \sin \beta_0 \cos \beta_3 e^{i\beta_4}, \quad h_{0,1} = \cos \beta_0 \cos \beta_1, \quad (\text{B13})$$

$$h_{1,0} = \sin \beta_0 \sin \beta_3 e^{i\beta_6}, \quad h_{1,1} = \cos \beta_0 \sin \beta_1 e^{i\beta_2}, \quad (\text{B14})$$

such that

$$a_{0,1} a_{1,0}^* \frac{(1 - e^{-i\theta})}{2} = c_1 \cos \alpha_0 \cos \alpha_1 \sin \alpha_0 \sin \alpha_3 e^{-i\alpha_6} + c_2 \cos \beta_0 \cos \beta_1 \sin \beta_0 \sin \beta_3 e^{-i\beta_6}, \quad (\text{B15a})$$

$$a_{0,0} a_{1,0}^* \frac{(1 - e^{-i\theta})}{2} = c_1 \cos \alpha_0 \sin \alpha_1 e^{i\alpha_2} \sin \alpha_0 \sin \alpha_3 e^{-i\alpha_6}, \quad (\text{B15b})$$

$$a_{1,1} a_{1,0}^* \frac{(1 - e^{-i\theta})}{2} = c_2 \cos \beta_0 \sin \beta_1 e^{i\beta_2} \sin \beta_0 \sin \beta_3 e^{-i\beta_6}, \quad (\text{B15c})$$

$$c_1 + c_2 = \sqrt{2} \sin \frac{\theta}{2}. \quad (\text{B15d})$$

For any $a_{j,k} \in \mathbb{C}$, we can always choose appropriate phase $\alpha_2, \beta_2, \alpha_6, \beta_6$ so that the phase of $a_{j,k}$ can be satisfied. Without loss of generality, we can only consider the case that $a_{j,k}$ are nonnegative real values,

$$a_{j,k} \geq 0 \quad \text{and} \quad \sum_{j,k} a_{j,k}^2 = 1. \quad (\text{B16})$$

Using (B10), we assume that $a_{1,0} \in (0, 1)$, and $a_{0,0} \in (0, 1)$ or $a_{1,1} \in (0, 1)$ here.

Now we show that (B15a)–(B15d) is viable by choosing appropriate parameters. We set $\alpha_2 = 0$ and $\alpha_j = \beta_j$, for $j = 0, 1, 2, 3, 6$. From (B15b)–(B15d), we assume

$$c_1 = \frac{\sqrt{2} \sin \frac{\theta}{2} a_{0,0}}{a_{0,0} + a_{1,1}}, \quad c_2 = \frac{\sqrt{2} \sin \frac{\theta}{2} a_{1,1}}{a_{0,0} + a_{1,1}}. \quad (\text{B17})$$

Using (B15b), (B15c), and (B17), Eq. (B15a) can be equivalently transformed into

$$a_{0,0} \cot \alpha_1 + a_{1,1} \cot \beta_1 = a_{0,1}. \quad (\text{B18})$$

It holds by choosing appropriate parameters $\alpha_1 = \beta_1$. Next, we show that Eqs. (B15b) and (B15c) can also be satisfied. First, we use (B18) to obtain that

$$\sin \alpha_1 = \sin \beta_1 = \frac{a_{0,0} + a_{1,1}}{\sqrt{a_{0,1}^2 + (a_{0,0} + a_{1,1})^2}}. \quad (\text{B19})$$

Then based on (B17) and (B19), we perform the transformation on (B15b) and (B15c) to put the free parameters $\alpha_0 = \beta_0$ and $\alpha_3 = \beta_3$ on the l.h.s. alone. They become the same equation as follows:

$$\cos \alpha_0 \sin \alpha_0 \sin \alpha_3 e^{-i\alpha_6} = e^{i(\frac{\pi}{2} - \frac{\theta}{2})} \frac{a_{1,0}}{\sqrt{2}} \sqrt{a_{0,1}^2 + (a_{0,0} + a_{1,1})^2}, \quad (\text{B20})$$

Note that $\frac{a_{1,0}}{\sqrt{2}} \sqrt{a_{0,1}^2 + (a_{0,0} + a_{1,1})^2} \in (0, \frac{1}{2}]$. So Eq. (B20) can always be satisfied for any $a_{j,k}$ in (B16) by choosing $\alpha_6 = \beta_6 = \frac{\theta}{2} - \frac{\pi}{2}$ and appropriate parameters $\alpha_0, \alpha_3, \beta_0, \beta_3$. Hence, (B15b) and (B15c) can be satisfied.

Based on (B5) and (B11)–(B14), the above analysis prove the existence of the $c^{(k)}$ and $F^{(k)}$ in (B5) and (B8) for any ρ . It means that there is a kind of decomposition following the rule in (B4), such that $c_1 + c_2 = \sqrt{2}$. Recall that the quantum Wasserstein distance between operations $\mathcal{D}(I, U_{CN} U_{CP} U_{CN})$ is defined by taking the minimization of $c_1 + c_2$ over all decompositions in (B4). One can obtain that $\mathcal{D}(I, U_{CN} U_{CP} U_{CN}) \leq \sqrt{2} \sin \frac{\theta}{2}$. Combining with (B3), it holds that $\mathcal{D}(I, U_\theta^{(3)}) = \mathcal{D}(I, U_{CN} U_{CP} U_{CN}) = \sqrt{2} \sin \frac{\theta}{2}$, which is the desired result. ■

By applying appropriate local unitary operation, the $U_\theta^{(k)}$'s can transform to each other, $(\sigma_x \otimes I) U_\theta^{(3)} (\sigma_x \otimes I) = U_\theta^{(1)}$, $(\sigma_x \otimes \sigma_x) U_\theta^{(3)} (\sigma_x \otimes \sigma_x) = U_\theta^{(2)}$, $(I \otimes \sigma_x) U_\theta^{(3)} (I \otimes \sigma_x) = U_\theta^{(4)}$. Since $\|\cdot\|_{W_1}$ is invariant under single-qubit unitary transformations, we obtain that $\mathcal{D}(I, U_\theta^{(k)}) = \sqrt{2} \sin \frac{\theta}{2}$, which is the result of Proposition 10.

2. The W_1 distance between I and single-qudit and permutation matrices

The proof of Proposition 8 is given here.

Proof. Using Lemma 2, we have $\mathcal{D}(U, V) = \mathcal{D}(I, VU^\dagger) = \max_{|\psi\rangle, |\phi\rangle \in \mathcal{S}_1} \sqrt{1 - |\langle \psi | VU^\dagger | \phi \rangle|^2}$, where the first equality holds from Property 7, and the second equality comes from $\| \alpha uu^* - \beta vv^* \|_1 = \sqrt{(\alpha + \beta)^2 - 4\alpha\beta} |\langle u, v \rangle|^2$.

First, we consider the case for $d = 2$, i.e., the operations act in the two-dimensional space. In order to obtain $\mathcal{D}(U, V)$, it suffices to compute the minimum of $|\langle \psi | VU^\dagger | \psi \rangle|$. Let $VU^\dagger = R^\dagger D_2 R$ be the spectral decomposition of VU^\dagger , where $D_2 = \text{diag}\{1, e^{i\alpha}\}$ and R is an unitary matrix. The state $|\xi\rangle := R|\psi\rangle$ is a single-qubit state. Suppose $|\xi\rangle = \cos \frac{\theta}{2} |0\rangle + e^{i\phi} \sin \frac{\theta}{2} |1\rangle$, for $0 \leq \theta \leq \pi$, $0 \leq \phi < 2\pi$. From $\min_{\theta, \phi} |\langle \xi | D_2 | \xi \rangle|^2 = \min_{\theta} |\cos^4 \frac{\theta}{2} + \sin^4 \frac{\theta}{2} + 2 \cos^2 \frac{\theta}{2} \sin^2 \frac{\theta}{2} \cos \alpha| = (1 + \cos \alpha)/2$, we have

$$\mathcal{D}(U, V) = \sqrt{1 - \min_{\theta, \phi} |\langle \xi | D_2 | \xi \rangle|^2} = \sqrt{\frac{1 - \cos \alpha}{2}}. \quad (\text{B21})$$

Then we consider the case for $d > 2$. As we have shown above, we assume that $VU^\dagger = R_d^\dagger D_d R_d$ is the spectral decomposition of VU^\dagger , where $D_d = \text{diag}\{e^{i\alpha_0}, e^{i\alpha_1}, \dots, e^{i\alpha_{d-1}}\}$ and R_d is a $d \times d$ unitary matrix. Suppose $|\psi\rangle$ is an arbitrary qudit state and $|\eta\rangle = R_d|\psi\rangle = \sum_j a_j |j\rangle$, for $\sum_j |a_j|^2 = 1$. We have

$$\begin{aligned} \mathcal{D}(U, V) &= \sqrt{1 - \min_{|\eta\rangle} |\langle \eta | D | \eta \rangle|^2} \\ &= \sqrt{1 - \min_{\sum_j |a_j|^2 = 1} \left| \sum_j |a_j|^2 e^{i\alpha_j} \right|^2}. \end{aligned} \quad (\text{B22})$$

The optimization is equivalent to minimizing the convex sum of the eigenvalues of VU^\dagger . ■

The proof of Proposition 12 is shown as follows.

Proof. Obviously, $\mathcal{D}(I, U_k) = 2$ can be obtained from $2 \geq \mathcal{D}(I, U_k) \geq \| |0, 1\rangle\langle 0, 1| - |1, 0\rangle\langle 1, 0| \|_{W_1} = 2$. From the local unitary invariance, we have $\|(U \otimes V)|0, 1\rangle\langle 0, 1| (U \otimes V)^\dagger -$

$(U \otimes V)|1, 0\rangle\langle 1, 0|(U \otimes V)^\dagger\|_{W_1} = 2$, where U, V are single qubit unitary operations. From Property 5, $\mathcal{D}(I, M) = 2$ is obtained. Next we show that the W_1 distance between identity and all order-4 permutation matrices can be derived. We consider the representation of order-4 permutation group $\{P_1, P_2, \dots, P_{24}\}$, where 15 permutation matrices are included in (6). By Proposition 12, one can obtain that the W_1 distance between every one of them and identity is equal to two. Other nine permutation matrices are listed below:

$$H_1 = I, H_2 = U_{CN}, \quad H_3 = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

$$H_4 = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \quad H_5 = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix},$$

$$H_6 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}, \quad H_7 = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

$$H_8 = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \quad H_9 = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{bmatrix}.$$

We analyze the distance for $\mathcal{D}(I, H_k)$. We have $(\sigma_x \otimes I)H_3(\sigma_x \otimes I) = H_2$, so $\mathcal{D}(I, H_2) = \mathcal{D}(I, H_3) = \sqrt{2}$.

The fact that $\mathcal{D}(I, H_4) = \mathcal{D}(I, H_5) = 1$ has been obtained in Proposition 14. Using the fact that $(I \otimes \sigma_x)H_6(I \otimes \sigma_x) = H_7$ and $U_{SW}H_6U_{SW} = H_2$, one has $\mathcal{D}(I, H_2) = \mathcal{D}(I, H_6) = \mathcal{D}(I, H_7) = \sqrt{2}$. By $(I \otimes \sigma_x)H_8(I \otimes \sigma_x) = H_9$, one has $\mathcal{D}(I, H_8) = \mathcal{D}(I, H_9)$. Using Lemma 3, we find that their lower bound is equal to two. Combined with Property 6, we have $\mathcal{D}(I, H_8) = \mathcal{D}(I, H_9) = 2$. We have now obtained the W_1 distance between identity and all the order-4 permutation matrices. ■

The proof of Proposition 14 is shown below.

Proof. We show the claim for $k = 1, 2$, and the claim for $k > 2$ can be obtained similarly.

First, we show that $\mathcal{D}(I^{\otimes n}, X \otimes I^{\otimes(n-1)}) = 1$ up to permutations of the qudits. For any pure states $\rho \in \mathcal{S}_n$ and $\sigma = (X \otimes I^{\otimes(n-1)})\rho(X \otimes I^{\otimes(n-1)})$, it holds that $\text{Tr}_1 \rho = \text{Tr}_1 \sigma$, i.e., ρ and σ are neighboring states. From Definition 2, the quantum W_1 distance assigns the distance at most one to any couple of neighboring states, so $\mathcal{D}(I^{\otimes n}, X \otimes I^{\otimes(n-1)}) \leq 1$. On the other hand, $\mathcal{D}(I^{\otimes n}, X \otimes I^{\otimes(n-1)}) = \max_\rho \|\rho - \sigma\|_{W_1} \geq \|\lvert 00 \dots 0 \rangle \langle 00 \dots 0 \rvert - \lvert 10 \dots 0 \rangle \langle 10 \dots 0 \rvert\|_{W_1} = 1$ by Lemma 5. Hence, $\mathcal{D}(I^{\otimes n}, X \otimes I^{\otimes(n-1)}) = 1$. Using the fact that $\|\cdot\|_{W_1}$ is invariant with respect to permutations of the qudits, $\mathcal{D}(I^{\otimes n}, I \otimes X \otimes I^{\otimes(n-2)}) = \dots = \mathcal{D}(I^{\otimes n}, I^{\otimes(n-1)} \otimes X) = 1$ is obtained.

Next we prove that $\mathcal{D}(I^{\otimes n}, X^{\otimes 2} \otimes I^{\otimes(n-2)}) = 2$ up to permutations of the qudits. From Definition 2, it holds that $\mathcal{D}(I^{\otimes n}, X^{\otimes 2} \otimes I^{\otimes(n-2)}) \geq \|\lvert 000 \dots 0 \rangle \langle 000 \dots 0 \rvert - \lvert 110 \dots 0 \rangle \langle 110 \dots 0 \rvert\|_{W_1} = 2$. By Property 10, we have $\mathcal{D}(I^{\otimes n}, X^{\otimes 2} \otimes I^{\otimes(n-2)}) \leq \mathcal{D}(I^{\otimes n}, X \otimes I^{\otimes(n-1)}) + \mathcal{D}(I^{\otimes n}, I \otimes X \otimes I^{\otimes(n-2)}) = 2$. The invariance under permutations can also be derived as the above case. ■

-
- [1] M. Cattaneo, M. A. C. Rossi, G. Garcia-Perez, R. Zambrini, and S. Maniscalco, Quantum simulation of dissipative collective effects on noisy quantum computers, *PRX Quantum* **4**, 010324 (2023).
 - [2] Y. Dong, S.-C. Zhang, Y. Zheng, H.-B. Lin, L.-K. Shan, X.-D. Chen, W. Zhu, G.-Z. Wang, G.-C. Guo, and F.-W. Sun, Experimental implementation of universal holonomic quantum computation on solid-state spins with optimal control, *Phys. Rev. Appl.* **16**, 024060 (2021).
 - [3] K. Beer, D. Bondarenko, T. Farrelly, T. J. Osborne, R. Salzman, D. Scheiermann, and R. Wolf, Training deep quantum neural networks, *Nat. Commun.* **11**, 808 (2020).
 - [4] K. Mitarai, M. Negoro, M. Kitagawa, and K. Fujii, Quantum circuit learning, *Phys. Rev. A* **98**, 032309 (2018).
 - [5] J. Dajka, J. Luczka, and P. Hanggi, Distance between quantum states in the presence of initial qubit-environment correlations: A comparative study, *Phys. Rev. A* **84**, 032120 (2011).
 - [6] N. Lashkari, Relative entropies in conformal field theory, *Phys. Rev. Lett.* **113**, 051602 (2014).
 - [7] G. De Palma, M. Marvian, D. Trevisan, and S. Lloyd, The quantum Wasserstein distance of order 1, *IEEE Trans. Inf. Theory* **67**, 6627 (2021).
 - [8] S. Friedland, M. Eckstein, S. Cole, and K. Zyczkowski, Quantum Monge-Kantorovich problem and transport distance between density matrices, *Phys. Rev. Lett.* **129**, 110402 (2022).
 - [9] G. Tóth and J. Pitrik, Quantum Wasserstein distance based on an optimization over separable states, *Quantum* **7**, 1143 (2023).
 - [10] E. Kashefi and A. Angrisani, Quantum local differential privacy and quantum statistical query model, *arXiv:2203.03591*.
 - [11] L. Li, K. Bu, D. E. Koh, A. Jaffe, and S. Lloyd, Wasserstein complexity of quantum circuits, *arXiv:2208.06306*.
 - [12] A. Acin, Statistical distinguishability between unitary operations, *Phys. Rev. Lett.* **87**, 177901 (2001).
 - [13] Y. Chen, H. Miyahara, L.-S. Bouchard, and V. Roychowdhury, Quantum approximation of normalized Schatten norms and applications to learning, *Phys. Rev. A* **106**, 052409 (2022).
 - [14] M. A. Nielsen, A simple formula for the average gate fidelity of a quantum dynamical operation, *Phys. Lett. A* **303**, 249 (2002).
 - [15] R. Duvenhage and M. Mapaya, Quantum wasserstein distance of order 1 between channels, *Infin. Dimens. Anal. Quantum. Probab. Relat. Top.* **26**, 2350006 (2023).
 - [16] C. A. Fuchs and J. van de Graaf, Cryptographic distinguishability measures for quantum-mechanical states, *IEEE Trans. Inf. Theory* **45**, 1216 (1999).
 - [17] E. Magesan, J. M. Gambetta, and J. Emerson, Scalable and robust randomized benchmarking of quantum processes, *Phys. Rev. Lett.* **106**, 180504 (2011).

- [18] Y. R. Sanders, J. J. Wallman, and B. C. Sanders, Bounding quantum gate error rate based on reported average fidelity, *New J. Phys.* **18**, 012002 (2015).
- [19] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, 2010).
- [20] J. Zhang, R. Zhang, N.-H. Chia, C.-N. Chou, Quantum meets the minimum circuit size problem, [arXiv:2108.03171](https://arxiv.org/abs/2108.03171).
- [21] M. A. Nielsen, M. R. Dowling, M. Gu, and, A. C. Doherty, Quantum computation as geometry, *Science* **311**, 1133 (2006).
- [22] M. A. Nielsen, A geometric approach to quantum circuit lower bounds, [arXiv:quant-ph/0502070](https://arxiv.org/abs/quant-ph/0502070).
- [23] D. Girolami and F. Anzà, Quantifying the difference between many-body quantum states, *Phys. Rev. Lett.* **126**, 170502 (2021).
- [24] X. Huang and L. Li, Query complexity of unitary operation discrimination, *Physica A* **604**, 127863 (2022).
- [25] C. K. Burrell, Geometry of generalized depolarizing channels, *Phys. Rev. A* **80**, 042330 (2009).
- [26] B. T. Kiani, G. De Palma, M. Marvian, Z. W. Liu, and S. Lloyd, Learning quantum data with the quantum earth mover's distance, *Quantum Sci. Technol.* **7**, 045002 (2022).
- [27] L. Chen and L. Yu, Entangling and assisted entangling power of bipartite unitary operations, *Phys. Rev. A* **94**, 022307 (2016).