Representation Matching For Remote Quantum Computing

Yuxiang Yang[1,2,*] and Masahito Hayashi[3,4,5,6,†]

[1] *Institute for Theoretical Physics, ETH Zürich, Switzerland*

[2] *QICI Quantum Information and Computation Initiative, Department of Computer Science, The University of Hong Kong, Pokfulam Road, Hong Kong*

[3] *Shenzhen Institute for Quantum Science and Engineering, Southern University of Science and Technology, Shenzhen 518055, China*

[4] *Guangdong Provincial Key Laboratory of Quantum Science and Engineering, Southern University of Science and Technology, Shenzhen 518055, China*

[5] *Shenzhen Key Laboratory of Quantum Science and Engineering, Southern University of Science and Technology, Shenzhen 518055, China*

[6] *Graduate School of Mathematics, Nagoya University, Nagoya 464-8602, Japan*

(Received 3 December 2020; accepted 21 April 2021; published 26 May 2021)

Many quantum computational tasks have inherent symmetries, suggesting a path to enhancing their efficiency and performance. Exploiting this observation, we propose representation matching, a generic probabilistic protocol for reducing the cost of quantum computation in a quantum network. We show that the representation-matching protocol is capable of reducing the communication or memory cost to almost the minimum in various tasks, including remote execution of unitary gate arrays, permutation gates, and unitary conjugation, as well as the storage and retrieval of unitary gates.

DOI: 10.1103/PRXQuantum.2.020327

I. INTRODUCTION

The past few years have witnessed tremendous progress in quantum computing and quantum communication. The union of technologies coming from these two directions will lead to a quantum internet [1], where remote nodes can exchange quantum data, execute protocols, and share computational power [2–5] via quantum communication channels [Fig. 1(a)].

One of the key issues for such a quantum network is the communication cost, quantified by the number of qubits needed to be sent via the communication channels. Any proposal for reducing the communication cost will have increasing importance, as the scale of quantum computation and quantum networks is expected to increase rapidly in the near future.

One possibility for reducing the communication cost is to consider probabilistic protocols. Many quantum protocols or algorithms, e.g., quantum key distribution [6], magic state distillation [7], and error mitigation [8], are probabilistic, where one repeats the protocol multiple times until it succeeds. Probabilistic protocols play a pivotal role in circumventing no-go theorems of quantum information [9–11]. Also, in discussing the classification of quantum complexity classes, e.g., the nondeterministic quantum polynomial-time complexity class, protocols with a small success probability play an important role (see Theorem 7 in Ref. [12], Theorem 1 in Ref. [13], and Ref. [14]). Here, in a network setting, the remote parties can communicate the desired computation using probabilistic protocols such as gate teleportation [15,16]. In communication scenarios, probabilistic protocols can overcome limitations, including restricted bandwidth and short memory life, and accomplish tasks that are impossible for deterministic protocols. Therefore, it may still be beneficial to employ a probabilistic protocol even if its expected communication cost, i.e., the expected amount of communication required until the protocol succeeds, is higher than that of deterministic protocols.

Another less explored observation is that many quantum computational tasks are associated with inherent symmetry. In other words, the relevant quantum gates form a group. For example, regarding computations on n qubits, there are the Pauli group $\mathrm{P}(n)$, the Clifford group $\mathrm{C}(n)$, the permutation group $\mathrm{S}(n)$, the braiding group $\mathrm{B}(n)$, and the special unitary group $\mathrm{SU}(2)$. Many fundamental tasks

[*] yangyu@ethz.ch

[†] hayashi@sustech.edu.cn

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

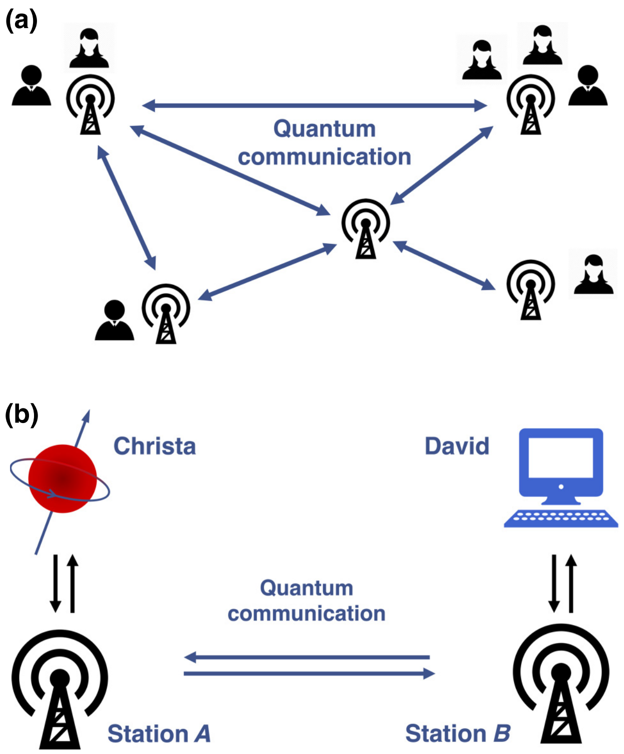

FIG. 1. (a) Quantum internet. In a quantum network, multiple parties are linked via quantum communication channels. (b) Remote quantum computation. The task is for David to apply a quantum gate to a quantum state, held by a remote party, Christa. The goal is to accomplish this objective while keeping the communication cost as low as possible.

in quantum information processing have group-theoretic structures, including cloning of states [17–20] and of gates [21–23], universal gate programming [24–30], storage and retrieval of unitary gates [31,32], and the inversion of general unitary operations [33]. A natural question is: Can we utilize the inherent symmetry of a task to enhance the performance and, in particular, to reduce the cost of communication in the remote setting?

In this work, we consider remote quantum computation in a network, where one party, David, would like to run a unitary gate on a quantum state held by Christa, who is far away from him [Fig. 1(b)]. We propose representation matching: a generic probabilistic protocol capable of reducing the communication cost of this task by exploiting the inherent symmetry of the task. We then prove a general lower bound on the communication cost of such a setting for remote quantum computing. When the computation is done with an array of n identical unitary gates of dimension d, our protocol attains the lower bound up to a small overhead that is independent of n. Moreover, the success probability of our protocol is much higher than that of a protocol based on gate teleportation. In particular, the ratio between the two success probabilities grows as n^{d^2-1}. We also apply our protocol to various tasks such as permutational quantum computing [34–38], unitary conjugation [33,39–41], and the storage and retrieval of gates [30–32].

The remaining part of the paper is organized as follows, In Sec. II, we introduce a few results and notations that are essential for our discussion. In Sec. III, we introduce the setting and the representation-matching protocol, and in Sec. IV we prove a general lower bound on the communication cost in the same setting. In Secs. V–VIII, we apply the representation-matching protocol to concrete quantum computational tasks. Finally, in Sec. IX, we conclude the paper with some discussions on future directions for research.

II. PRELIMINARY CONSIDERATIONS

For a Hilbert space $\mathcal{H}$ and a vector $|\psi\rangle \in \mathcal{H}$, we use the notation $\psi := |\psi\rangle\langle\psi|$ to denote the projector on the one-dimensional subspace spanned by $|\psi\rangle$. The space of linear operators from a Hilbert space $\mathcal{H}$ to another Hilbert space $\mathcal{K}$ is denoted by $L(\mathcal{H},\mathcal{K})$. When the two Hilbert spaces coincide, we use the shorthand $L(\mathcal{H}) := L(\mathcal{H},\mathcal{H})$. In this paper, we focus on finite-dimensional quantum systems, with $\dim(\mathcal{H}) < \infty$. For a quantum system with a Hilbert space $\mathcal{H}$, the set of quantum states is denoted by $\mathcal{S}(\mathcal{H}) := \{\rho \in L(\mathcal{H})|\mathrm{Tr}[\rho] = 1, \langle\psi|\rho|\psi\rangle \geq 0\ \forall|\psi\rangle \in \mathcal{H}\}$.

A quantum process that deterministically transforms an input system into a (possibly different) output system is called a *quantum channel*. A quantum channel transforming an input system with a Hilbert space $\mathcal{H}^{\mathrm{in}}$ into an output system with a (possibly different) Hilbert space $\mathcal{H}^{\mathrm{out}}$ is a completely positive trace-preserving map $\mathcal{C} : L(\mathcal{H}^{\mathrm{in}}) \to L(\mathcal{H}^{\mathrm{out}})$. A probabilistic quantum process transforming an input system with a Hilbert space $\mathcal{H}_{\mathrm{in}}$ into an output system with a (possibly different) Hilbert space $\mathcal{H}^{\mathrm{out}}$ is called a *quantum operation* and is described by a completely positive trace-nonincreasing map $\mathcal{M} : L(\mathcal{H}^{\mathrm{in}}) \to L(\mathcal{H}^{\mathrm{out}})$. A quantum operation takes a quantum state in $\mathcal{H}^{\mathrm{in}}$ as input and produces a subnormalized state in $\mathcal{H}^{\mathrm{out}}$ as output.

We frequently consider the Hilbert space $\mathcal{H}^{\otimes n}$, i.e., the Hilbert space of n identical systems, each with a Hilbert space $\mathcal{H}$. We treat it using a basic knowledge of representation theory, and we refer interested readers to textbooks, e.g., Ref. [42] or Chapter 6 of Ref. [43], for more information. Here we introduce a few useful results without further explanation.

The structure of $\mathcal{H}^{\otimes n}$ is characterized by the Schur-Weyl duality, which states that there exists an isometry transforming $\mathcal{H}^{\otimes n}$ into block diagonal form:

$$\mathcal{H}^{\otimes n} \simeq \bigoplus_{\lambda\in\mathcal{R}_n} \mathcal{H}^{\lambda} \otimes \mathcal{M}^{\lambda}, \tag{1}$$

where $\mathcal{R}_n$ is the collection of all Young diagrams of n boxes; $\mathcal{H}^{\lambda}$ is the irreducible-representation subspace of

$\mathrm{SU}(d)$, the special unitary group of degree d, characterized by the Young diagram λ; and $\mathcal{M}^\lambda$ is the irreducible-representation subspace of $\mathrm{S}(n)$, the symmetric group of degree n. The isometry, called the Schur transform, can be implemented efficiently on a quantum computer [44–46]. Since the irreducible representations are in one-to-one correspondence with the Young diagrams, the set $\mathcal{R}_n$ is the collection of all Young diagrams with (at most) d rows and n boxes, defined as

$$\mathcal{R}_n := \left\{ \lambda = (\lambda_1, \ldots, \lambda_d) : \lambda_i \in \mathbb{N}, \right.$$
$$\left. \lambda_i \geq \lambda_j, \forall i, j; \sum_{i=1}^{d} \lambda_i = n \right\}. \tag{2}$$

Finally, we introduce a few dimensional factors that will be useful. The first is the total number of irreducible representations in the decomposition in Eq. (1), i.e., the cardinality of the set $\mathcal{R}_n$. By the definition of $\mathcal{R}_n$ [Eq. (2)], we have the following bound:

$$|\mathcal{R}_n| \leq (n+1)^{d-1}. \tag{3}$$

Next, the dimension of an $\mathrm{SU}(d)$ irreducible representation λ can be obtained via the following formula:

$$d_\lambda = \frac{\prod_{1\leq i<j\leq d}(\lambda_i - \lambda_j - i + j)}{\prod_{k=1}^{d-1} k!}. \tag{4}$$

The dual of d_λ, the dimension m_λ of an $\mathrm{S}(n)$ irreducible representation λ, has the following expression (see, e.g., Ref. [42]):

$$m_\lambda = \frac{n! \prod_{1\leq j<k\leq d}(\lambda_j - \lambda_k - j + k)}{\prod_{i=1}^{d}(\lambda_i + d - i)}. \tag{5}$$

Denoting by d_{R} the maximum of d_λ over $\lambda \in \mathcal{R}_n$, from the above formula we have [43, Eq. (6.16)]

$$d_{\mathrm{R}} := \max_{\lambda \in \mathcal{R}_n} d_\lambda \leq (n+1)^{[d(d-1)]/2}. \tag{6}$$

We denote by d_{tot} the sum of all d_λ in $\mathcal{R}_n$,

$$d_{\mathrm{tot}} := \sum_{\lambda \in \mathcal{R}_n} d_\lambda, \tag{7}$$

which can be bounded as

$$d_{\mathrm{tot}} \leq d_{\mathrm{R}} |\mathcal{R}_n| \leq (n+1)^{[(d+2)(d-1)]/2}. \tag{8}$$

Finally, we denote by $d_{\mathrm{tot,sq}}$ the sum of the squared dimensions of all irreducible representations,

$$d_{\mathrm{tot,sq}} := \sum_{\lambda \in \mathcal{R}_n} d_\lambda^2 = \binom{n + d^2 - 1}{n}, \tag{9}$$

having used [47, Eq. (57)]. It follows that

$$\log d_{\mathrm{tot,sq}} = (d^2 - 1) \log n + O(1), \tag{10}$$

where $O(1)$ denotes a term that does not depend on n.

III. REPRESENTATION-MATCHING PROTOCOL

Consider a common remote quantum computing scenario as shown in Fig. 1(b). The task is for David to execute a target computation U^{target} on a state ψ^{in}, held by another remote party, Christa. That is, the final output should be $U^{\mathrm{target}} \psi^{\mathrm{in}} (U^{\mathrm{target}})^\dagger$, located on Christa's side. The goal is to design a protocol for the two stations A and B, which provide the data transmission service for Christa and David, that reduces their total communication cost. The setting is *blind*, which means that the stations do not know U^{target} or ψ^{in} *a priori*. This is also the case in most practical applications, since the users of a communication link would usually require their information to be kept private.

When the target computation is a unitary representation of a group G on a fixed Hilbert space, we can express it as

$$U_g^{\mathrm{target}} = V^\dagger \left(\sum_{r \in \mathcal{R}} |r\rangle\langle r|_{\mathrm{I}} \otimes (U_g^r)_{\mathrm{R}} \otimes (I_{m_r})_{\mathrm{M}} \right) V, \quad g \in \mathsf{G}. \tag{11}$$

Here U^r is an irreducible representation of G indexed by r, V is a (g-independent) unitary gate, $\{|r\rangle\}$ is an orthonormal basis for indices of the irreducible representations, and m_r denotes the multiplicity of the irreducible representation U^r in the decomposition of U. In Eq. (11), the first register is referred to as the *index register*, I, the second as the *representation register*, R, and the third as the *multiplicity register*, M. For instance, for $g \in \mathrm{SU}(2)$ we have $U_g^{\otimes n} \simeq \sum_{j=0}^{n/2} |j\rangle\langle j| \otimes U_g^j \otimes I_{m_j}$ (assuming for simplicity that n is even), where each irreducible representation U^j is characterized by a spin number j. We discuss special unitary groups more in later sections.

To fulfill the computational task, a straightforward approach is to communicate both the index register and the representation register, and, as U_g^{target} acts trivially on it, the multiplicity register can be stored locally on Christa's side. The overall transmission cost, in terms of qubits, is thus twice of the cost of transmitting the index register and the representation register. By merging these two registers into one (see Step 4 of Protocol 1 later), the cost can be reduced to

$$c_{\max} = 2 \lceil \log d_{\mathrm{tot}} \rceil, \tag{12}$$

where

$$d_{\mathrm{tot}} := \sum_{r \in \mathcal{R}} d_r \tag{13}$$

Protocol 1 Representation matching protocol.

Input: An arbitrary input state $|\psi^{\rm in}\rangle$ with $V|\psi^{\rm in}\rangle \in \mathcal{H}^{\rm I}\otimes\mathcal{H}^{\rm R}\otimes\mathcal{R}$ on Christa's side, where V is defined by Eq. (11).

Output: The quantum state $U_g^{\rm target}|\psi^{\rm in}\rangle$ at Christa's side with some probability; the successful case is heralded.

1: Station A gets an input state, applies V, sends only the representation register R and stores locally the index register I.

2: Station B prepares the ansatz register A in the state:

$$|f^{\rm ans}\rangle_{\rm A} := \frac{1}{\sqrt{|\mathcal{R}|}}\sum_{r\in\mathcal{R}}|r\rangle_{\rm A}. \tag{14}$$

3: Station B performs $V^\dagger$ [cf. Eq. (11)] on A, R, and an ancillary multiplicity register in a trivial state, asks David to perform $U_g^{\rm target}$ and then performs V.

4: Station B sends the output state back to station A. Notice that, although station B has to send out two registers A and R, the required communication is not necessarily the size of the two registers. Explicitly, since the state to be sent lives in a subspace of $\mathcal{H}^{\rm A}\otimes\mathcal{H}^{\rm R}$, there is an isometry $W : |r\rangle_{\rm A}\otimes|m\rangle_{\rm R} \to |l\rangle_{\rm C}$ encoding both registers into a memory, where $\{|m\rangle\}_{m=1}^{\max d_r}$ is a basis of the representation register and $\{|l\rangle\}_{l=0}^{\infty}$ is the Fock basis. The part where the state has no support can be cut off and the remaining part of the memory only has dimension $d_{\rm tot}$.

5: **Coherent matching test.** Station A restores both registers from the memory by applying $W^\dagger$ and then performs the coherent matching test, which is a quantum operation $\{\mathcal{M}_{\rm yes},\mathcal{M}_{\rm no}\}$, jointly on the ansatz register A and the index register I, where

$$\mathcal{M}_{\rm yes}(\cdot) := M_{\rm yes}(\cdot)M_{\rm yes}^\dagger \tag{15}$$

$$M_{\rm yes} := \sum_r \langle r|_{\rm A}\otimes|r\rangle\langle r|_{\rm I}. \tag{16}$$

6: Station A applies $V^\dagger$ and returns the state to Christa if the measurement outcome is "yes"; otherwise restart from Step 1.

and $\lceil\cdot\rceil$ denotes the ceiling function [48]. In the decomposition in Eq. (11), the index register I contains no information on the desired computation g. Meanwhile, it is usually rather costly to transmit the register I. For example, the transmission cost of the index register for $U_g^{\otimes n}$ [$g\in {\rm SU}(2), n\gg 1$], i.e., multiple parallel uses of a qubit gate, is nearly half of the total transmission cost. Nevertheless, for any input state living in more than one irreducible representation, the index register is indispensable.

Here we propose a probabilistic protocol for this task. Our protocol reduces the cost by executing the computation on the representation register based on an ansatz of r and postselecting the case where the ansatz holds. In this way, the communication cost can be reduced (see Fig. 2).

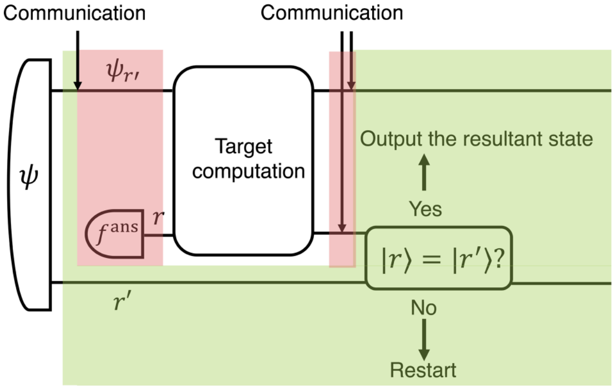

FIG. 2. Procedure of the representation-matching protocol. The part in green describes the action of station A, and the part in red describes the action of station B. Station A sends only the representation register and stores the index register locally. Station B prepares an ansatz state $|f^{\rm ans}\rangle$ and executes the target computation jointly on it and the representation register received from A. Station B then returns the output to A, which performs a coherent matching test to see if the target computation has been performed correctly. In this way, the cost of transmitting the index register can be avoided.

In Protocol 1, the communication cost consists of two parts: the cost of sending R, which has dimension

$$d_{\rm R} := \max_{r\in\mathcal{R}} d_r, \tag{17}$$

to station B and the cost of sending both A and R, which have dimension $d_{\rm tot}$ in total, back to station A. Since the index register is transmitted only once, the communication cost of the representation-matching protocol is

$$c_{\rm rm} = \lceil\log d_{\rm R}\rceil + \lceil\log d_{\rm tot}\rceil, \tag{18}$$

where $d_{\rm R}$ is defined by Eq. (17) and $d_{\rm tot}$ is defined by Eq. (13). Compared with Eq. (12), the representation-matching protocol achieves a reduction of

$$\Delta c := c_{\max} - c_{\rm rm} = \lceil\log d_{\rm tot}\rceil - \lceil\log d_{\rm R}\rceil \tag{19}$$

qubits. The price of the reduction is a risk of failure. The success probability of the representation-matching protocol can be straightforwardly evaluated as

$$p_{\rm rm} = \frac{1}{|\mathcal{R}|}. \tag{20}$$

There may be some scenarios where the input state $\psi^{\rm in}$ from Christa contains valuable information, e.g., outcomes of previous computations. As a result, Christa might not know $\psi^{\rm in}$ and cannot reprepare it arbitrarily. Since the representation protocol is probabilistic, it seems that $\psi^{\rm in}$ will be lost or corrupted if the protocol fails. Given this concern, Christa may want to avoid corruption of ψ^{in}. However, in the following we show that, if we modify the coherent matching test slightly, we can avoid corrupting ψ^{in} even when the representation protocol fails. The modification is based on the following observation: the stations extract no information on ψ^{in} by doing the coherent matching test, because the success probability in Eq. (20) and the measurement outcome are independent of it (even though the postmeasurement state does depend on ψ^{in}).

To modify the coherent matching test, we now specify the quantum operation $\mathcal{M}_{\mathrm{no}}$ for the failure case. We assume an additive group structure for $\mathcal{R}$ and define the unitary gate $V_{\hat{r}} := \sum_{r\in\mathcal{R}} |r+\hat{r}\rangle\langle r|_{\mathrm{I}}$ on $\mathcal{H}^{\mathrm{I}}$. The set of measurement outcomes for the modified coherent matching test is given as $\mathcal{R}$. The measurement operation $\mathcal{M}_{\hat{r}}$ corresponding to the outcome $\hat{r}\in\mathcal{R}$ is the following:

$$\mathcal{M}_{\hat{r}}(\cdot) := M_{\hat{r}}(\cdot)M_{\hat{r}}^{\dagger}, \tag{21}$$

$$M_{\hat{r}} := \sum_{r} \langle r+\hat{r}|_{\mathrm{A}} \otimes |r\rangle\langle r|_{\mathrm{I}}. \tag{22}$$

The outcome $\hat{r}=0$ corresponds to "yes," and other outcomes correspond to "no." The protocol continues even if the outcome is "no." When the outcome $\hat{r}$ is observed, the resultant state is $V_{-\hat{r}}VU_g^{\mathrm{target}}V^{\dagger}V_{\hat{r}}V|\psi^{\mathrm{in}}\rangle$. Hence, if we apply the unitary gate $(V_{-\hat{r}}VU_g^{\mathrm{target}}V^{\dagger}V_{\hat{r}}V)^{-1}$, we can recover the original state ψ^{in}, which can be considered as a special case of the quantum rewinding lemma [49, Lemma 8].

When the outcome $\hat{r}$ is not zero, i.e., when the protocol fails, the parties can perform another round and apply the representation protocol with the target unitary gate $U_g^{\mathrm{target}}(V^{\dagger}V_{-\hat{r}}VU_g^{\mathrm{target}}V^{\dagger}V_{\hat{r}}V)^{-1} = U_g^{\mathrm{target}}V^{\dagger}V_{-\hat{r}}V(U_g^{\mathrm{target}})^{\dagger}V^{\dagger}V_{\hat{r}}V$ to the above resultant state, i.e., David applies the unitary gate $U_g^{\mathrm{target}}V^{\dagger}V_{-\hat{r}}V(U_g^{\mathrm{target}})^{\dagger}V^{\dagger}V_{\hat{r}}V$ in the second round. When the outcome of the coherent matching test in the second round is 0, the resultant state is the desired state. Therefore, it is possible to repeat our protocol for multiple rounds until it succeeds. The probability that the protocol succeeds within n rounds is $1-(1-p_{\mathrm{rm}})^n$, and thus we have the following result.

Remark 1. *The success probability of Protocol 1 can be amplified to* $1-\epsilon$ *for arbitrarily small* ϵ*. This requires executing the protocol for* $O(\log(\epsilon)/\log(1-p_{\mathrm{rm}})$ *rounds with target gates* U_g^{target}*,* $(U_g^{\mathrm{target}})^{\dagger}$*, and* V_r*, where* p_{rm} *is given by Eq. (20).*

In addition to the blind setting, Protocol 1 works in another setting, the so-called *visible* setting, in which station B has access to the target group element $g\in\mathrm{G}$ and can execute any arbitrary g-dependent computation. Remember that station B is given only black-box access to U_g^{target} in the blind setting. Apparently, the visible setting may require a lower communication cost, as the constraint is more relaxed. In the following, we show a lower bound on the communication cost in the visible setting, and the lower bound is asymptotically achieved by Protocol 1 even in the blind setting.

IV. LOWER BOUND ON THE COMMUNICATION COST

Here we prove a lower bound on the communication cost of remote quantum computing, where one party performs a computation U_g^{target} for a state held by another faraway party. We consider *any arbitrary* deterministic or probabilistic protocol, i.e., we show a bound on the required communication cost that holds no matter how small the success probability is. We also consider the visible setting, where, in contrast to the previous setting, station B and David are now considered as one single party. Notice that any bound for the visible setting also holds for the blind setting, and thus the lower bound can be used to evaluate the performance of representation matching.

In the visible setting, the action of station A is still required to be independent of the input state, whereas the action of station B (together with David) can be described by a quantum operation $\mathcal{S}_g : L(\mathcal{H}^{\mathrm{M,\,in}}) \to L(\mathcal{H}^{\mathrm{M,\,out}})$ acting on a memory system, as illustrated in Fig. 3. On the side of station A, an encoder is first performed to package (part of) the input state, and a decoder is performed after receiving the state from station B. The encoder is given as an isometric quantum channel $\mathcal{E} : L(\mathcal{H}^{\mathrm{tot}}) \to L(\mathcal{H}^{\mathrm{M,\,in}} \otimes \mathcal{H}^{\mathrm{M}})$. The decoder, on the other hand, is given as a quantum operation $\mathcal{D} : L(\mathcal{H}^{\mathrm{M,\,out}} \otimes \mathcal{H}^{\mathrm{M}}) \to L(\mathcal{H}^{\mathrm{tot}})$. Here $\mathcal{H}^{\mathrm{M}}$ denotes the Hilbert space of a local memory. Notice that $\mathcal{E}$ is assumed to be isometric without loss of generality, because any postselection or partial trace can be postponed to $\mathcal{D}$. The dimension of $\mathcal{H}^{\mathrm{M},x}$ ($x=$ in, out) is denoted by $d_{\mathrm{M},x}$.

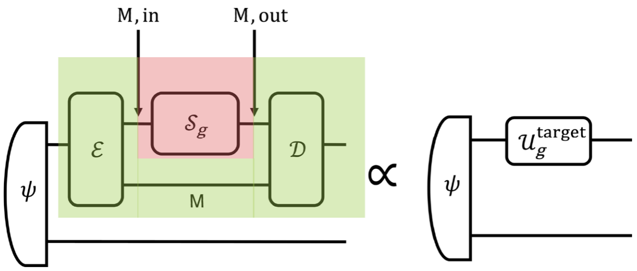

FIG. 3. The visible setting of remote quantum computing. In the visible setting of gate compression, the goal is still to implement U_g^{target} on an input ψ held by Christa. The only difference from the blind setting is that station B knows which gate to perform or, equivalently, the value of g. Therefore, station B is allowed to perform a generic g-dependent quantum operation, which we denote by $\mathcal{S}_g$.

Here the overall communication cost is $c := \lceil \log d_{\mathrm{M,\,in}} \rceil + \lceil \log d_{\mathrm{M,\,out}} \rceil$, regarding which we have the following result.

Theorem 1. *Consider probabilistic remote computing of a target gate U_g^{target} as defined by Eq. (11). The total communication cost in the visible setting is lower bounded as*

$$c \geq c_{\min} := \lceil \log d_{\mathrm{tot,sq}} \rceil \tag{23}$$

regardless of the success probability, where

$$d_{\mathrm{tot,sq}} := \sum_{r \in \mathcal{R}} d_r^2. \tag{24}$$

The proof can be found in Appendix A. We stress that Theorem 1 is the ultimate limit for *all protocols*. That is, no matter how small a success probability we allow, the lower bound always holds.

By Theorem 1, the representation-matching protocol (see Protocol 1) has at most an overhead of

$$\delta c := c_{\mathrm{rm}} - c_{\min} = \lceil \log d_{\mathrm{R}} \rceil + \lceil \log d_{\mathrm{tot}} \rceil - \lceil \log d_{\mathrm{tot,sq}} \rceil. \tag{25}$$

As we show soon, in many concrete applications the overhead δc is only a few qubits, and thus Protocol 1 is almost optimal in terms of communication efficiency.

V. COMPRESSION OF GATE ARRAYS

In the following, we demonstrate several concrete applications of Protocol 1. The first task we consider is the compression of an array of n identical d-dimensional unitary gates, i.e., $U_g^{\otimes n}$ with $g \in \mathrm{SU}(d)$. By the Schur-Weyl duality [Eq. (1)], there exists a unitary gate (the Schur transform) U_{Sch} transforming $U_g^{\otimes n}$ into block diagonal form:

$$U_{\mathrm{Sch}} U_g^{\otimes n} U_{\mathrm{Sch}}^{\dagger} = \sum_{\lambda \in \mathcal{R}_n} |\lambda\rangle\langle\lambda|_{\mathrm{I}} \otimes \left(U_g^{\lambda}\right)_{\mathrm{R}} \otimes \left(I_{m_\lambda}\right)_{\mathrm{M}}, \tag{26}$$

where U_g^{λ} is now the $\mathrm{SU}(d)$ irreducible representation characterized by the Young diagram λ and m_λ is the dimension of the $\mathrm{S}(n)$ representation, now serving as the multiplicity of U_g^{λ}.

In the remote computing setting, this amounts to saying that David would like to run n parallel uses of U_g on a remote state of Christa. Such a setting is also frequently encountered in quantum sensor networks [50,51], where the preparation of the sensor state and the application of the unknown unitary gates happen at different locations. The objective is to communicate $U_g^{\otimes n}$ with a lower cost, i.e., to "compress" the gate array. Protocol 1 can be readily applied to fulfill the task, with U_g^{target} [see Eq. (11)] being $U_g^{\otimes n}$ for $g \in \mathrm{SU}(d)$.

Next, we discuss the performance of Protocol 1. Using Eq. (20), the probability of success is

$$p_{\mathrm{rm}} = \frac{1}{|\mathcal{R}_n|}. \tag{27}$$

Therefore, by using the bound in Eq. (3), we have

$$p_{\mathrm{rm}} \geq \left(\frac{1}{n+1}\right)^{d-1}. \tag{28}$$

Meanwhile, combining Eq. (18) with Eqs. (6) and (8), the communication cost of Protocol 1 is upper bounded by

$$c_{\mathrm{rm}} \leq (d^2-1)\log(n+1) + 2, \tag{29}$$

which, according to Theorem 1, attains the optimal scaling with n. Let us now examine the communication-cost saving Δc, defined by Eq. (19), which equals the gap between c_{rm} and $c_{\max}$ (the communication cost without compression). Observing that $d_{\mathrm{tot}} \geq d_{\mathrm{tot,sq}}/d_{\mathrm{R}}$, we have

$$\Delta c \geq \log\left(\frac{d_{\mathrm{tot}}}{d_{\mathrm{R}}}\right) - 1 \tag{30}$$

$$\geq \log\left(\frac{d_{\mathrm{tot,sq}}}{2 d_{\mathrm{R}}^2}\right) \tag{31}$$

$$\geq \log\left(\frac{\binom{n+d^2-1}{d^2-1}}{2(n+1)^{d(d-1)}}\right), \tag{32}$$

having used Eqs. (6) and (9) in the last step. In the large-n asymptotics, the above implies that

$$\Delta c \geq (d-1)\log n - O(1). \tag{33}$$

On the other hand, employing Eqs. (3) and (8), we have

$$\Delta c \leq \log\left(\frac{d_{\mathrm{tot}}}{d_{\mathrm{R}}}\right) + 1 \tag{34}$$

$$\leq \log|\mathcal{R}_n| + 1 \tag{35}$$

$$\leq (d-1)\log(n+1) + 1. \tag{36}$$

Summarizing the above inequalities, we identify the scaling of Δc as

$$\Delta c = (d-1)\log n + O(1). \tag{37}$$

Next we show the asymptotic optimality of Protocol 1. According to Eq. (10), the gap between the cost of Protocol

1 and the lower bound $c_{\min}$ is at most

$$\delta c = O(1) \tag{38}$$

qubits. In particular, consider the qubit case, where $g \in \mathrm{SU}(2)$. Assuming first that n is even, the lower bound can be explicitly evaluated as

$$\begin{aligned} c_{\min} &= \left\lceil \log\left(\sum_{j=0}^{n/2}(2j+1)^2\right)\right\rceil \\ &\geq \log\left(\frac{1}{6}(n+1)(n+2)(n+3)\right). \end{aligned} \tag{39}$$

On the other hand, the cost of Protocol 1 is

$$\begin{aligned} c_{\mathrm{rm}} &= \lceil \log{(n+1)} \rceil + \left\lceil \log\left(\sum_{j=0}^{n/2}(2j+1)\right)\right\rceil \\ &\leq \log\left(\frac{1}{4}(n+1)(n+2)^2\right) + 2. \end{aligned} \tag{40}$$

The overhead is

$$\delta c = c_{\mathrm{rm}} - c_{\min} \leq \log\left(\frac{6(n+2)}{(n+3)}\right) < \log 6. \tag{41}$$

Similarly, for n odd one can also show that $\delta < \log 6$. Therefore the overhead is no more than 2 qubits. In Fig. 4, we numerically compare the communication cost c_{rm} with $c_{\min}$ and $c_{\max}$, from which one can observe that the asymptotic optimality of Protocol 1 matches the above discussion.

We summarize the performance in the following theorem.

Theorem 2. *Protocol 1 fulfills the task of compressing $U_g^{\otimes n}$ [$g \in SU(d)$] perfectly. The total communication cost is given by Eq. (29) and attains the optimal scaling with n. The success probability is given by Eq. (28) and scales as $n^{-(d-1)}$.*

We can compare the performance of Protocol 1 with that of another protocol, which is based on the gate teleportation approach [15,16]. One can retrieve a d-dimensional unitary gate U_g from the maximally entangled state $|\Phi_g^+\rangle := (U_g \otimes I)|\Phi^+\rangle$ [$|\Phi_+\rangle := \sum_i (1/\sqrt{d})|i\rangle \otimes |i\rangle$] by performing a generalized Bell test jointly on part of it and an arbitrary input state $|\psi\rangle$:

$$(I \otimes B_j)(|\Phi_g^+\rangle \otimes |\psi\rangle) = \frac{1}{d}(U_g\sigma_j|\psi\rangle) \otimes (I \otimes \sigma_j)|\Phi^+\rangle \tag{42}$$

for $j = 0, \ldots, d^2 - 1$, where $B_j := (I \otimes \sigma_j)\Phi^+(I \otimes \sigma_j)$ and σ_j is a generalized Pauli operator. In particular, $\sigma_0 := I$ is the identity. Therefore, with probability $(1/d)^2$, we get the outcome $j = 0$ and $U_g|\psi\rangle$ as desired. A naive approach to compressing gate arrays runs as follows:

(1) Station B lets David apply each of the n unitary gates in the array to $|\Phi^+\rangle$ and sends the resultant state, which is $|\Phi_g^+\rangle^{\otimes n}$, to station A.

(2) Station A performs gate teleportation locally with the received state and the input from Christa.

Evidently, this approach fares much worse than the representation-matching protocol in terms of both the communication cost and the success probability. Indeed, the

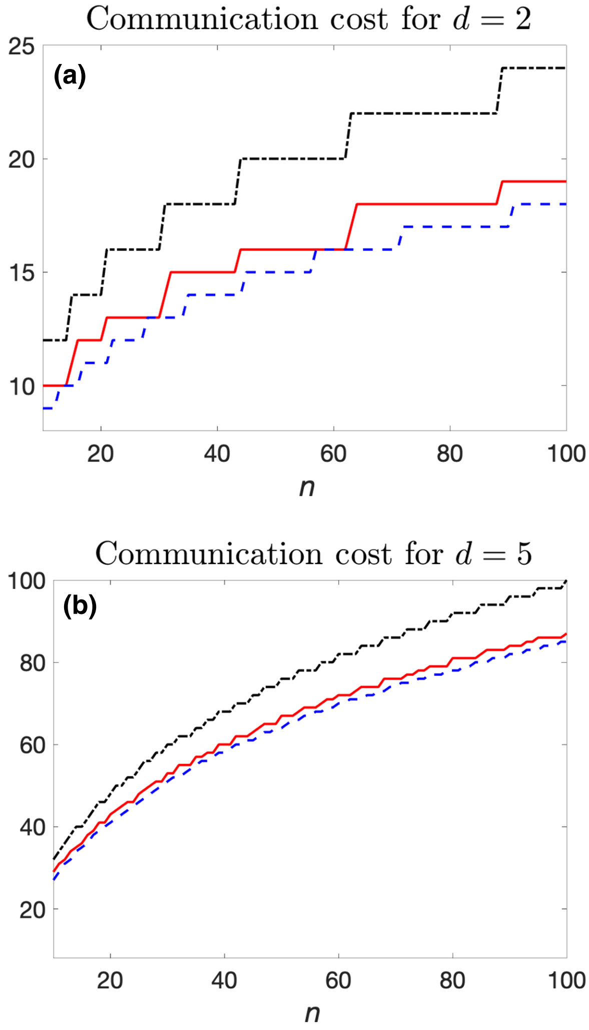

FIG. 4. Communication cost of Protocol 1. The communication cost of compressing $U_g^{\otimes n}$ [$g \in \mathrm{SU}(d)$] is plotted as a function of n for (a) $d = 2$ and (b) $d = 5$. In both plots, the red solid lines represent c_{rm}, the cost of representation matching (Protocol 1); the blue dashed lines represent $c_{\min}$, the lower bound on the cost given by Theorem 1; and the black dash-dotted lines represent $c_{\max}$, the original cost. Notice that Protocol 1 is asymptotically optimal: for $n = 100$, the overhead $\delta c := c_{\mathrm{rm}} - c_{\min}$ is only 1 for $d = 2$ and 2 for $d = 5$.

Protocol 2 Gate teleportation based compression of $U_g^{\otimes n}$.

1: Station B applies $U_{\rm Sch}$, asks David to perform $U_g^{\otimes n}$, and performs $U_{\rm Sch}^{\dagger}$ in sequential order on registers A, $\rm R_1$, and M of the state

$$|\Phi_n^+\rangle_{\rm AR_1R_2M} := \sum_{\lambda\in\mathcal{R}_n}\sqrt{\frac{d_\lambda}{d_{\rm tot}}}|\lambda\rangle_{\rm A}\otimes|\Phi_\lambda^+\rangle_{\rm R_1R_2}\otimes|\eta_0\rangle_{\rm M}, \tag{43}$$

where $|\Phi_\lambda^+\rangle$ is the maximally entangled state on $\mathcal{H}_\lambda\otimes\mathcal{H}_\lambda$ and $|\eta_0\rangle$ is a fixed (but otherwise arbitrary) state. Station B then sends out the resultant state $|\Phi_{g,n}\rangle$ to station A.

2: Station A performs a generalised Bell measurement $\{B_j\}_{j=0}^{d_{\rm tot}^2-1}$ ($B_0 = |\Phi_n^+\rangle\langle\Phi_n^+|$) jointly on part of $|\Phi_{g,n}\rangle$ and the input state from Christa. The protocol is successful if and only if the Bell measurement yields $j=0$.

communication cost is $n\lceil\log(d^2)\rceil$, which is exponentially higher, and the success probability is $(1/d)^{2n}$, which vanishes exponentially with n.

Instead, we can consider an improved version of the gate-teleportation-based approach by exploiting the decomposition in Eq. (26) (Protocol 2). The communication cost of the teleportation-based protocol is $c_{\rm min}$, which is very close to that of Protocol 1 according to Eq. (38). However, as the success probability of gate teleportation is inversely proportional to the square of the system dimension, the success probability of Protocol 2 is only

$$p_{\rm tele} = \frac{1}{d_{\rm tot}^2}, \tag{44}$$

where $d_{\rm tot}$ is defined by Eq. (7). In contrast, Protocol 1 has a much higher probability of success. The ratio between the two success probabilities is

$$\frac{p_{\rm rm}}{p_{\rm tele}} = O\left(n^{d^2-1}\right). \tag{45}$$

The advantage of Protocol 1 is obvious even in the nonasymptotic regime, as illustrated in Fig. 5. In conclusion, our representation-matching protocol outperforms even the improved version of gate teleportation in the task of gate-array compression.

VI. COMPRESSION OF PERMUTATION GATES

Quantum computation consisting only of permutations of subsystems has gained increasing interest, for it offers new insight into topological quantum computing [34–38]. In the following, we show that Protocol 1 applies not only to SU(d) but also to finite groups such as the permutation group of n particles S(n).

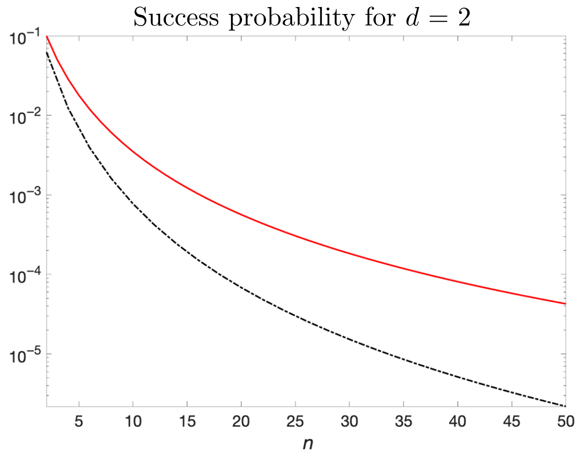

FIG. 5. Success probability of Protocols 1 and 2. The success probability for compressing $U_g^{\otimes n}$ [$g\in$ SU(d)] is plotted as a function of n for $d=2$. The red solid line represents $p_{\rm rm}$, the success probability of representation matching (Protocol 1), while the black dash-dotted line represents $p_{\rm tele}$, the success probability of gate teleportation (Protocol 2).

Consider all permutations of n qudits. By the Schur-Weyl duality [see Eq. (1)], we can decompose a unitary permutation gate V_g^n with $g\in$ S(n) as

$$U_{\rm Sch}V_g^nU_{\rm Sch}^{\dagger} = \sum_{\lambda\in\mathcal{R}_n}|\lambda\rangle\langle\lambda|_{\rm I}\otimes(I_{d_\lambda})_{\rm M}\otimes(V_g^{\lambda,n})_{\rm R}, \tag{46}$$

where $U_{\rm Sch}$ is the Schur transform, $\mathcal{R}_n$ is defined by Eq. (2), $V_g^{\lambda,n}$ is the irreducible representation associated with the Young diagram λ, and d_λ is the dimension of the SU(d) irreducible representation U^λ, now serving as the multiplicity of $V_g^{\lambda,n}$. In particular, the dimension of $V_g^{\lambda,n}$, m_λ, is given by Eq. (5).

Since m_λ is usually quite large, even the minimum communication cost, given by Theorem 1, grows linearly with n instead of with $\log n$. Nevertheless, Protocol 1 is still capable of reducing the communication cost by a moderate amount. Here we examine the quantity Δc, defined by Eq. (19), which is the communication-cost saving achieved by Protocol 1. In Fig. 6, one can see that Δc grows as a function of n. In the meantime, the gap δc between the cost of Protocol 1 and the minimum cost (see Theorem 1) remains very low (less than 3 qubits).

We can also characterize the scaling of Δc analytically. Consider the case of qubits. The dimension of each spin-j irreducible representation [Eq. (5)] reduces to

$$m_j = \frac{2j+1}{n+1}\binom{n+1}{\frac{n}{2}-j}. \tag{47}$$

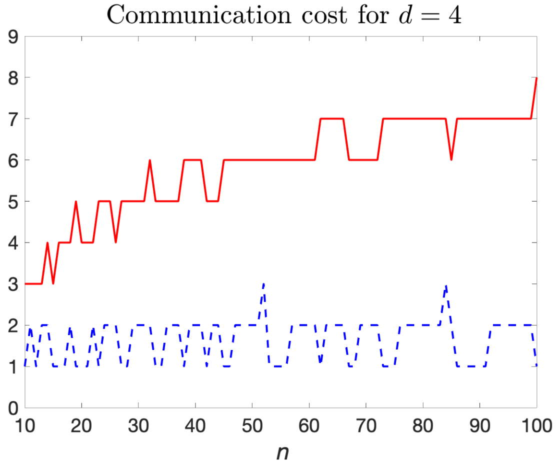

FIG. 6. Cost reduction of Protocol 1 for permutation gates. The red solid line represents the communication-cost saving of Protocol 1 [see Eq. (19)] for the task of compressing permutation gates of n qudits for $d = 4$. The blue dashed line represents the gap δc [see Eq. (25)].

For large n, the maximum of m_j is achieved with $j \sim \sqrt{n}$. On the other hand, one can verify that

$$d_{\mathrm{tot}} = \sum_{j=0}^{n/2} m_j = \frac{n+2}{n+1}\binom{n+1}{\frac{n}{2}}, \tag{48}$$

having assumed n even for simplicity. By Eq. (19), we have

$$\Delta c \geq \log\left(\frac{d_{\mathrm{tot}}}{d_{\mathrm{R}}}\right) - 1 \tag{49}$$

$$\geq \frac{1}{2}\log n - O(1). \tag{50}$$

Therefore, the communication-cost saving of Protocol 1 goes to infinity with n. In particular, comparing Eq. (50) with Eq. (37), we conclude that, for large n, the communication-cost saving achieved by Protocol 1 for permutation gates is at least half of the saving for unitary gate arrays, when $d = 2$.

VII. CONJUGATION OF UNITARY GATES

Besides compression, Protocol 1 also fits the task of conjugating quantum gates remotely. Consider the same remote quantum computing setting as in previous sections, but suppose that the goal now is to execute n parallel uses of the gate U_g^* for an unknown $g \in \mathrm{SU}(d)$ on an arbitrary input state of Christa, with David performing m parallel uses of U_g. Here U^* denotes the complex conjugate of U.

From representation theory, it is known that the complex conjugate of an irreducible representation λ of $\mathrm{SU}(d)$ is isomorphic to the irreducible representation with Young diagram $\bar{\lambda}$. Here $\bar{\lambda}$ is the Young diagram *associated to* λ, which is obtained by changing the box number of each column from k to $d-k$. That is, for every λ, there exists a unitary gate V^{λ} such that

$$(U^{\lambda})^* = V^{\lambda} U^{\bar{\lambda}} (V^{\lambda})^{\dagger}. \tag{51}$$

For example, the two-dimensional irreducible representation of SU(2) satisfies $U^* = \sigma_y U \sigma_y$, where σ_y is the Pauli y matrix.

In order to apply our approach, we need the computation to include $U^{\bar{\lambda}}$ for every $\lambda \in \mathrm{SU}(d)$. In particular, the $\bar{\lambda}$ associated to $\lambda = (n, 0, \ldots, 0)$ has $(d-1)n$ boxes. It is also straightforward that all other associated Young diagrams have more boxes. Therefore, we need $U_g^{\otimes m}$ with $m = (d-1)n$ to implement the computation $(U_g^*)^{\otimes n}$.

The computation can be implemented remotely and probabilistically using Protocol 1, where in Step 3 station B performs (by querying U_g m times from David)

$$U = V U_{\mathrm{Sch}} U_g^{\otimes m} U_{\mathrm{Sch}}^{\dagger} V^{\dagger}, \tag{52}$$

$$V := \sum_{\lambda \in \mathcal{R}_n} |\lambda\rangle\langle\lambda| \otimes V^{\lambda}, \tag{53}$$

with V^{λ} defined by Eq. (51) and $m = (d-1)n$. The communication cost and the success probability are exactly the same as in the case of unitary-gate-array compression, and the (asymptotic) optimality can be shown in the same way using Theorem 1. Therefore we have the following result.

Theorem 3. *Protocol 1 fulfills the task of the gate conversion $U_g^{\otimes m} \to (U_g^*)^{\otimes n}$ perfectly for every $m \geq (d-1)n$. The total communication cost is given by Eq. (29) and attains the optimal scaling with n. The success probability is given by Eq. (28) and scales as $n^{-(d-1)}$.*

VIII. STORAGE AND RETRIEVAL OF GATE ARRAYS

The idea of representation matching applies not only to quantum networks over space, which we have considered in the previous sections, but also to quantum networks over time. Here, as a typical example, we apply representation matching to the task of storage and retrieval of unitary gate arrays [32]. As shown in Fig. 7, the goal is to store m instances of an unknown unitary gate U_g and to retrieve n ($\leq m$) instances later on. In comparison with remote quantum computing, the task is for David to apply a computation to a quantum state held by Christa, who is located at a future location of David. This task is also known as quantum learning [31,52] and has recently been shown to be closely related to the optimal programming of quantum gates [30].

Here we employ the idea of representation matching and propose Protocol 3 for the storage and retrieval of gate arrays.

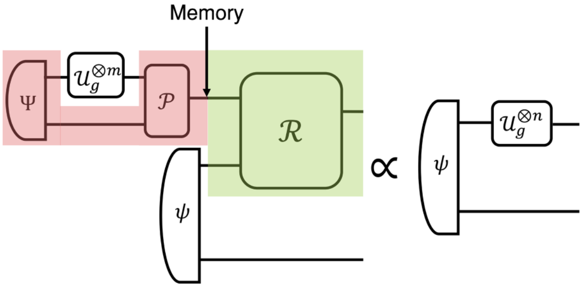

FIG. 7. The task of storage and retrieval of unitary gates. The task is separated into two stages: storage (in red) and retrieval (in green). In the storage stage, $U_g^{\otimes m}$ [$g \in \mathrm{SU}(d)$] is applied to (part of) a state Ψ of the memory, followed by postprocessing, $\mathcal{P}$. In the retrieval stage, a retrieval operation $\mathcal{R}$ is applied to extract $U_g^{\otimes n}$ from the memory and apply it to an arbitrary input state. The whole process is probabilistic, and (heralded) failure is allowed.

Protocol 3 Probabilistic storage and retrieval of $U_g^{\otimes n}$.

1: (Storage.) Apply $(U_{\rm Sch} U_g^{\otimes n} U_{\rm Sch}^\dagger)_{\rm AR_1M} \otimes I_{\rm R_2}$ [see Eq. (26)] on the memory state

$$|\Psi_n\rangle_{\rm AR_1R_2} := \sum_{\lambda\in\mathcal{R}_n} \frac{d_\lambda}{\sqrt{d_{\rm tot,sq}}} |\lambda\rangle_A \otimes |\Phi_\lambda^+\rangle_{\rm R_1R_2} \otimes |\eta_0\rangle_{\rm M}, \quad (54)$$

where $|\eta_0\rangle$ is an arbitrary fixed state. Store the resultant state into a memory register.

2: (Retrieval.) To apply the stored unitary on any input state, which can be cast into the form

$$|\psi\rangle = \sum_{\lambda\in\mathcal{R}_n} c_\lambda |\lambda\rangle_{\rm I} \otimes |\psi_\lambda\rangle_{\rm RP} \quad (55)$$

where P is a purification register and $\sum_{\lambda\in\mathcal{R}_n} |c_\lambda|^2 = 1$. perform a quantum operation $\{\mathcal{N}_{\rm yes}, \mathcal{N}_{\rm no}\}$ jointly on A, I, R and R_2. The successful operation is defined as

$$\mathcal{N}_{\rm yes}(\cdot) := N_{\rm yes}(\cdot) N_{\rm yes}^\dagger \quad (56)$$

$$N_{\rm yes} := \sum_{\lambda\in\mathcal{R}_n} \langle\lambda|_{\rm I_1} \otimes |\lambda\rangle\langle\lambda|_{\rm I} \otimes \langle\Phi_\lambda^+|_{\rm RR_2}. \quad (57)$$

Next, we analyze the performance of our protocol. The input state $|\psi\rangle$ [see Eq. (1)] and the memory state that stores $U_g^{\otimes n}$ can be jointly expressed as

$$|\psi'\rangle = \sum_{\lambda',\lambda\in\mathcal{R}_n} \frac{c_\lambda d_{\lambda'}}{\sqrt{d_{\rm tot,sq}}} |\lambda'\rangle_{\rm A} \otimes |\lambda\rangle_{\rm I} \otimes |\psi_\lambda\rangle_{\rm RP} \otimes |\Phi_{g,\lambda'}^+\rangle_{\rm R_1R_2}. \quad (58)$$

The role of the quantum operation $\mathcal{N}_{\rm yes}$ is twofold. First, it checks whether A and I are in the same state; second, if so, it performs a gate teleportation, extracting $U_{g,\lambda}$ from R_1 and applying it to R. Therefore, Protocol 3 integrates our representation-matching idea into the gate-teleportation approach. The protocol succeeds if and only if both the representation matching and the gate teleportation are successfully performed, and then the state becomes

$$|\psi_{\rm out}\rangle = \sum_{\lambda\in\mathcal{R}_n} c_\lambda |\lambda\rangle_{\rm I} \otimes (U_g^\lambda \otimes I_{\rm P}) |\psi_\lambda\rangle_{\rm RP}, \quad (59)$$

which is exactly as desired. Therefore, the protocol has no error.

For the gate-retrieval-and-storage protocol, the probability of success is given by

$$p_{\rm rs} = \mathrm{Tr}\left[\mathcal{N}_{\rm yes} \otimes \mathcal{I}_{\rm P\,R_1}(\psi)\right] = \frac{1}{d_{\rm tot,sq}}. \quad (60)$$

Meanwhile, the protocol requires a quantum memory of size

$$c_{\rm rs} = \lceil \log d_{\rm tot,sq} \rceil = (d^2-1)\log n + O(1), \quad (61)$$

where we have used Eq. (10). On the other hand, a lower bound on the required memory size can be determined using Theorem 1. Indeed, notice that a storage-and-retrieval protocol can always be employed to fulfill a remote computation. To do so, station B stores the gate in a memory and sends it to station A, which retrieves it from the memory later; the communication cost of such a protocol is equal to the size of the memory. Therefore, the lower bound in Theorem 1 also applies here to the storage-and-retrieval task. By comparing Eq. (61) and Theorem 1, we find that

$$c_{\rm rs} = c_{\rm min}. \quad (62)$$

Therefore, Protocol 3 is optimal in terms of memory efficiency for the task of storage and retrieval of unitary gates. Summarizing, we have the following theorem.

Theorem 4. *Protocol 3 fulfills the task of storage and retrieval of $U_g^{\otimes n}$ [$g \in SU(d)$] perfectly. The memory cost is given by Eq. (61) and is optimal. The success probability is given by Eq. (60) and scales as $n^{-(d^2-1)}$.*

Our result extends that of Ref. [32], where the $n \to 1$ task (i.e., storing n uses and retrieving one use) was considered. We conclude this section by remarking that adaptations of Protocol 3 can be applied in other tasks. For instance, one can consider the problem of reversing quantum computation: the gate conversion $U_g^{\otimes n} \to (U_g^\dagger)^{\otimes n}$. To this end, notice that $U^\dagger = (U^T)^*$, where U^* and U^T are the complex conjugate and the transpose, respectively, of U. We thus divide the gate-reversal task into two separate steps: $U \to U^*$ and $U^* \to (U^*)^T$. The former can be accomplished using the techniques described in Sec. VII, while the latter can be achieved with a variant of Protocol 3.

IX. CONCLUSION AND DISCUSSION

We have studied how to reduce the communication or memory cost for certain types of computational tasks in the quantum internet. Our main contribution is twofold. For one thing, derive a lower bound on the cost. For another, we propose representation matching, a generic probabilistic protocol capable of asymptotically achieving our lower bound in many practical scenarios. In addition, the success probability of representation matching is also much higher than that of protocols based on existing ideas, e.g., gate teleportation. Compared with existing protocols for remote quantum computing, e.g., Ref. [53,54], our protocols make more use of the specific structure of the problem (i.e., by implementing multiple uses of a gate drawn from a group) to achieve higher cost reduction.

From the previous examples, one can see that, while the communication-cost reduction grows with n, the success probability also vanishes. One may ask if there are certain values of n for which even the *average* cost of representation matching, i.e., $c_{\mathrm{rm}}/p_{\mathrm{rm}}$, can be lower than that of a deterministic protocol. The answer is negative: indeed, according to Eqs. (12), (18), and (20), the average cost of representation matching cannot be lower than $c_{\max}$ if $|\mathcal{R}| \geq 2$. Nevertheless, there are still good reasons, both practically and conceptually, to consider probabilistic protocols such as representation matching. For example, representation matching serves as an excellent means of *deterrence*. Imagine, in the same setting as that shown in Fig. 1(a), that Christa is now an authority who needs to ensure that David is being honest and is performing the desired computation. She could do this by sending a state to David and letting him run some quantum gates on it. If David is caught cheating, he has to pay an extremely heavy fine. For such a task, it is more favorable to use representation matching rather than to use deterministic protocols: even if the protocol has a chance of failure, in which case Christa may fail to detect David's dishonesty, David will not risk it if the fine is high enough. Notice that the coherent matching test (see Protocol 1) is performed *after* David returns the output state, so he cannot predict whether the protocol will succeed (or whether Christa will decide to perform the coherent matching test at all). Representation matching could also be used if the bandwidth of the communication channel is limited and deterministic protocols are prohibited by this limitation. Finally, it is always fundamentally meaningful to explore the ultimate limits of quantum information theory, even at the cost of a small success probability. Instances of such research include quantum cloning [20], quantum metrology [55], quantum programming [32], and, here in this work, reduction of the cost of remote quantum computing.

In this work, we have focused on zero-error protocols, but our idea of representation matching can be extended to the approximate setting, which could be an interesting direction for future research. In particular, the communication cost of deterministic and approximate remote execution of unitary gate arrays has been shown to be closely related to quantum metrology [56], which raises the intriguing question of whether a similar phenomenon exists in the probabilistic setting.

The key of point of our protocol is the zero-error property. When no error is allowed at all and the communication cost is huge, our method is useful. To increase the success probability under the zero-error condition, we may need to increase the amount of quantum communication. In this work, we have not studied how much quantum communication is needed to achieve a certain success probability under the zero-error condition. It seems that this trade-off requires a new technical tool. Therefore, it is an interesting future problem to study the trade-off between the success probability and the amount of quantum communication under the zero-error condition.

ACKNOWLEDGMENTS

We thank Tomoyuki Morimae for helpful comments. YY was supported by the Swiss National Science Foundation via the National Center for Competence in Research "QSIT" and via Project No. 200020_165843, by the ETH Pauli Center for Theoretical Studies, and by the AFOSR via Grant No. FA9550-19-1-0202. MH was supported in part by the Guangdong Provincial Key Laboratory (Grant No. 2019B121203002).

APPENDIX: LOWER BOUND ON THE COMMUNICATION COST OF ZERO-ERROR REMOTE COMPUTATION

Given a group G, the task under consideration is to perform U_g^{target}, which is a projective unitary representation on $\mathcal{H}^{\mathrm{tot}}$ for any $g \in \mathsf{G}$, on a remote state. This representation contains the irreducible representations $\{U_g^r\}_{r\in\mathcal{R}}$, i.e.,

$$U_g^{\mathrm{target}} = \bigoplus_{r\in\mathcal{R}} U_g^r, \tag{A1}$$

where the irreducible-representation space of U_g^r is $\mathcal{H}^r$, with dimension d_r.

We prepare two lemmas before proving the main result.

Lemma 1. *The dimension of the linear subspace spanned by $\{\langle u_{l'}|U_g^{\mathrm{target}}|u_l\rangle\}_{l,l'}$ as a function space over G is $\sum_{r\in\mathcal{R}} d_r^2$, where $|u_l\rangle$ is a basis of $\mathcal{H}^{\mathrm{tot}}$.*

Proof. This is due to the following basic property of irreducible representations: the matrix elements $(U_g^r)_{i,j}$ for different irreducible representations r are linearly independent, and therefore the total dimension is $\sum_{r\in\mathcal{R}} d_r^2$. ■

Lemma 2. *Assume that two linear maps U and V from $\mathcal{H}_1$ to $\mathcal{H}_2$ satisfy*

$$U|\psi\rangle = c_\psi V|\psi\rangle, \tag{A2}$$

and also that the kernel of U is $\{0\}$. Then, c_ψ does not depend on ψ.

Proof. Assume that $|\psi_1\rangle, |\psi_2\rangle$ are linearly independent. Considering their superposition, we have

$$U(|\psi_1\rangle + |\psi_2\rangle) = c_{\psi_1+\psi_2} V(|\psi_1\rangle + |\psi_2\rangle) = c_{\psi_1} V|\psi_1\rangle + c_{\psi_2} V|\psi_2\rangle. \tag{A3}$$

By assumption, $V|\psi_1\rangle$ and $V|\psi_2\rangle$ are also independent. Therefore, the above equalities imply $c_{\psi_1} = c_{\psi_2} = c_{\psi_1+\psi_2}$. ■

We consider the less stringent *visible* setting, i.e., when station B knows what U_g^{target} is. In the visible setting, the action of station B can be described by a quantum operation $\mathcal{S}_g : L(\mathcal{H}^{\text{M, in}}) \to L(\mathcal{H}^{\text{M, out}})$ acting on a memory system, as illustrated in Fig. 3. As for station A, an encoder is first performed to package (part of) the input state, and a decoder is performed after receiving the state from station B. The encoder is given as an isometric quantum channel $\mathcal{E} : L(\mathcal{H}^{\text{M, in}}) \to L(\mathcal{H}^{\text{tot}} \otimes \mathcal{H}^{\text{M}})$. The decoder, on the other hand, is given as a quantum operation $\mathcal{D} : L(\mathcal{H}^{\text{M, out}} \otimes \mathcal{H}_{\text{R}}) \to L(\mathcal{H}^{\text{tot}})$. Notice that $\mathcal{E}$ is assumed to be isometric without loss of generality, because any postselection or partial trace can be postponed to $\mathcal{D}$. The dimension of $\mathcal{H}^{\text{M},x}$ ($x = \text{in, out}$) is denoted by $d_{\text{M},x}$.

Now, we impose the perfect-recovery condition, which reads

$$U_g^{\text{target}} \rho (U_g^{\text{target}})^\dagger = c_{g,\rho} \mathcal{D} \circ \mathcal{S}_g \circ \mathcal{E}(\rho) \tag{A4}$$

for any $g \in \text{G}$ and $\rho \in \mathcal{S}(\mathcal{H}^{\text{M, in}})$. Here $c_{g,\rho}$ is the reciprocal of the success probability of the postselection.

Theorem 5. *Under the condition of Eq. (A4), we have*

$$d_{\text{M,in}} d_{\text{M,out}} \geq \sum_{r \in \mathcal{R}} d_r^2. \tag{A5}$$

Proof. First, we invoke Stinespring dilations for all the channels (operations) involved. We choose the environment systems $\mathcal{H}_{\text{E}_1}$ and $\mathcal{H}_{\text{E}_2}$ and the postselection systems $\mathcal{H}_{\text{S}_1}$ and $\mathcal{H}_{\text{S}_2}$. Station B's operation can be purified as

$$\mathcal{S}_g(\cdot) = \text{Tr}_{\text{E}_1}\left[\langle\psi_{\text{S}_1}|V_g(\cdot)V_g^\dagger|\psi_{\text{S}_1}\rangle\right] \tag{A6}$$

for an isometry $V_g : \mathcal{H}^{\text{M, in}} \to \mathcal{H}^{\text{M, out}} \otimes \mathcal{H}^{\text{E}_1} \otimes \mathcal{H}^{\text{S}_1}$ and a pure state $|\psi_{\text{S}_1}\rangle$ on $\mathcal{H}^{\text{S}_1}$. Similarly, for the decoder we have

$$\mathcal{D}(\cdot) = \text{Tr}_{\text{E}_2}\left[\langle\psi_{\text{S}_2}|V_{\mathcal{D}}(\cdot)V_{\mathcal{D}}^\dagger|\psi_{\text{S}_2}\rangle\right], \tag{A7}$$

with $V_{\mathcal{D}} : \mathcal{H}^{\text{M, out}} \otimes \mathcal{H}^{\text{M}} \to \mathcal{H}^{\text{tot}} \otimes \mathcal{H}^{\text{E}_2} \otimes \mathcal{H}^{\text{S}_2}$ being an isometry and $|\psi_{\text{S}_2}\rangle$ being a pure state on $\mathcal{H}^{\text{S}_2}$. Moreover, by definition, the encoder is of the form

$$\mathcal{E}(\cdot) = V_{\mathcal{E}}(\cdot)V_{\mathcal{E}}^\dagger, \tag{A8}$$

with $V_{\mathcal{E}}$ being an isometry.

With the above dilations, the condition of Eq. (A4) can be rewritten as

$$U_g^{\text{target}} \rho (U_g^{\text{target}})^\dagger = c_{g,\rho} \text{Tr}_{\text{E}_1,\text{E}_2} \times \left(\langle\psi_{\text{S}_1},\psi_{\text{S}_2}|V_{\mathcal{D}}V_gV_{\mathcal{E}}\rho\left(V_{\mathcal{E}}V_gV_{\mathcal{D}}\right)^\dagger|\psi_{\text{S}_1},\psi_{\text{S}_2}\rangle\right), \tag{A9}$$

with a coefficient $c_{g,\rho}$ for $g \in \text{G}$ and $\rho \in \mathcal{S}(\mathcal{H}^{\text{M, in}})$.

Now, we choose a basis on $\{|e_k\rangle\}$ on $\mathcal{H}^{\text{E}_1}$ and a basis on $\{|f_j\rangle\}$ on $\mathcal{H}^{\text{E}_2}$. Then, we have

$$U_g^{\text{target}} \rho (U_g^{\text{target}})^\dagger = c_{g,\rho} \sum_{i,j} \langle e_i, f_j, \psi_{\text{S}_1}, \psi_{\text{S}_2}|V_{\mathcal{D}}V_gV_{\mathcal{E}}\rho \left(V_{\mathcal{E}}V_gV_{\mathcal{D}}\right)^\dagger|e_i, f_j, \psi_{\text{S}_1}, \psi_{\text{S}_2}\rangle. \tag{A10}$$

Since $\rho = |\psi\rangle\langle\psi|$ is a pure state, for i,j we have

$$U_g^{\text{target}}|\psi\rangle\langle\psi|(U_g^{\text{target}})^\dagger = c_{g,\psi,i,j}\langle e_i, f_j, \psi_{\text{S}_1}, \psi_{\text{S}_2}|V_{\mathcal{D}}V_gV_{\mathcal{E}}|\psi\rangle \langle\psi|\left(V_{\mathcal{E}}V_gV_{\mathcal{D}}\right)^\dagger|e_i, f_j, \psi_{\text{S}_1}, \psi_{\text{S}_2}\rangle, \tag{A11}$$

with coefficients $\{c_{g,\psi,i,j}\}$. We define the map $V_{g,i} := \langle e_i, \psi_{\text{S}_1}|V_g$ from $\mathcal{H}^{\text{M, in}}$ to $\mathcal{H}^{\text{M, out}}$ and the map $V_{\mathcal{D},j} := \langle f_j, \psi_{\text{S}_2}|V_{\mathcal{D}}$ from $\mathcal{H}^{\text{M, out}} \otimes \mathcal{H}_{\text{M}}$ to $\mathcal{H}^{\text{tot}}$, which are both linear. Substituting into Eq. (A11), we have

$$U_g^{\text{target}}|\psi\rangle\langle\psi|(U_g^{\text{target}})^\dagger = c_{g,\psi,i,j} V_{\mathcal{D},j}V_{g,i}V_{\mathcal{E}}|\psi\rangle\langle\psi|(V_{\mathcal{D},j}V_{g,i}V_{\mathcal{E}})^\dagger. \tag{A12}$$

Thus, there exists a $\theta_{g,\psi,i,j}$ such that

$$U_g^{\text{target}}|\psi\rangle = \sqrt{c_{g,\psi,i,j}}\, e^{i\theta_{g,\psi,i,j}} V_{\mathcal{D},j}V_{g,i}V_{\mathcal{E}}|\psi\rangle. \tag{A13}$$

Because of Lemma 2, $\sqrt{c_{g,\psi,i,j}}e^{i\theta_{g,\psi,i,j}}$ is independent of $|\psi\rangle$, and so we rename it as $\alpha_{g,i,j}$. Thus, we have

$$U_g^{\text{target}} = \alpha_{g,i,j}V_{\mathcal{D},j}V_{g,i}V_{\mathcal{E}}. \tag{A14}$$

Defining the matrix $V_{g,i,j} := \alpha_{g,i,j}V_{g,i}$ from the space $\mathcal{H}_{\text{M, in}}$ to the space $\mathcal{H}_{\text{M, out}}$, we have

$$U_g^{\text{target}} = V_{\mathcal{D},j}V_{g,i,j}V_{\mathcal{E}}. \tag{A15}$$

On the one hand, the dimension of the linear space spanned by $\{\langle u_{l'}|V_{g,i,j}|u_l\rangle\}_{l,l'}$ is at most $d_{\text{M, in}}d_{\text{M, out}}$. On the other hand, since U_g^{target} can be obtained by a linear transformation on $V_{g,i,j}$, we know that the dimension of $\text{Span}\{\langle u_{l'}|V_{g,i,j}|u_l\rangle\}_{l,l'}$ is lower bounded by the dimension of $\text{Span}\{\langle u_{l'}|U_g^{\text{target}}|u_l\rangle\}_{l,l'}$. Applying Lemma 1, we get Eq. (A5). ■

[1] H. J. Kimble, The quantum internet, Nature **453**, 1023 (2008).

[2] A. M. Childs, Secure assisted quantum computation, Quantum Inf. Comput. **5**, 456 (2005).

[3] A. Broadbent, J. Fitzsimons, and E. Kashefi, in *2009 50th Annual IEEE Symposium on Foundations of Computer Science* (IEEE, 2009), p. 517.

[4] J. F. Fitzsimons and E. Kashefi, Unconditionally verifiable blind quantum computation, Phys. Rev. A **96**, 012303 (2017).

[5] E. Kashefi and A. Pappa, Multiparty delegated quantum computing, Cryptography **1**, 12 (2017).

[6] G. Brassard, in *the IEEE International Conference on Computers, Systems and Signal Processing (Bangalore, India)* (1984), p. 175.

[7] S. Bravyi and A. Kitaev, Universal quantum computation with ideal Clifford gates and noisy ancillas, Phys. Rev. A **71**, 022316 (2005).

[8] K. Temme, S. Bravyi, and J. M. Gambetta, Error Mitigation for Short-Depth Quantum Circuits, Phys. Rev. Lett. **119**, 180509 (2017).

[9] W. K. Wootters and W. H. Zurek, A single quantum cannot be cloned, Nature **299**, 802 (1982).

[10] M. A. Nielsen and I. L. Chuang, Programmable Quantum Gate Arrays, Phys. Rev. Lett. **79**, 321 (1997).

[11] V. Bužek, M. Hillery, and R. Werner, Optimal manipulations with qubits: Universal-NOT gate, Phys. Rev. A **60**, R2626 (1999).

[12] K. Fujii, H. Kobayashi, T. Morimae, H. Nishimura, S. Tamate, and S. Tani, arXiv:1509.07276.

[13] K. Fujii, H. Kobayashi, T. Morimae, H. Nishimura, S. Tamate, and S. Tani, Impossibility of Classically Simulating One-Clean-Qubit Model with Multiplicative Error, Phys. Rev. Lett. **120**, 200502 (2018).

[14] M. J. Bremner, R. Jozsa, and D. J. Shepherd, Classical simulation of commuting quantum computations implies collapse of the polynomial hierarchy, Proc. R. Soc. A: Math. Phys. Eng. Sci. **467**, 459 (2011).

[15] D. Gottesman and I. L. Chuang, Demonstrating the viability of universal quantum computation using teleportation and single-qubit operations, Nature **402**, 390 (1999).

[16] S. D. Bartlett and W. J. Munro, Quantum Teleportation of Optical Quantum Gates, Phys. Rev. Lett. **90**, 117901 (2003).

[17] N. Gisin and S. Massar, Optimal Quantum Cloning Machines, Phys. Rev. Lett. **79**, 2153 (1997).

[18] D. Bruss, A. Ekert, and C. Macchiavello, Optimal Universal Quantum Cloning and State Estimation, Phys. Rev. Lett. **81**, 2598 (1998).

[19] D. Bruß, M. Cinchetti, G. M. D'Ariano, and C. Macchiavello, Phase-covariant quantum cloning, Phys. Rev. A **62**, 012302 (2000).

[20] G. Chiribella, Y. Yang, and A. C.-C. Yao, Quantum replication at the Heisenberg limit, Nat. Commun. **4**, 1 (2013).

[21] G. Chiribella, G. M. D'Ariano, and P. Perinotti, Optimal Cloning of Unitary Transformation, Phys. Rev. Lett. **101**, 180504 (2008).

[22] W. Dür, P. Sekatski, and M. Skotiniotis, Deterministic Superreplication of One-Parameter Unitary Transformations, Phys. Rev. Lett. **114**, 120503 (2015).

[23] G. Chiribella, Y. Yang, and C. Huang, Universal Superreplication of Unitary Gates, Phys. Rev. Lett. **114**, 120504 (2015).

[24] J. Kim, Y. Cheong, J.-S. Lee, and S. Lee, Storing unitary operators in quantum states, Phys. Rev. A **65**, 012302 (2001).

[25] G. Vidal, L. Masanes, and J. I. Cirac, Storing Quantum Dynamics in Quantum States: A Stochastic Programmable Gate, Phys. Rev. Lett. **88**, 047905 (2002).

[26] M. Hillery, V. Bužek, and M. Ziman, Probabilistic implementation of universal quantum processors, Phys. Rev. A **65**, 022301 (2002).

[27] A. Brazier, V. Bužek, and P. L. Knight, Probabilistic programmable quantum processors with multiple copies of program states, Phys. Rev. A **71**, 032306 (2005).

[28] S. Ishizaka and T. Hiroshima, Asymptotic Teleportation Scheme as a Universal Programmable Quantum Processor, Phys. Rev. Lett. **101**, 240501 (2008).

[29] A. M. Kubicki, C. Palazuelos, and D. Pérez-García, Resource Quantification for the No-Programing Theorem, Phys. Rev. Lett. **122**, 080505 (2019).

[30] Y. Yang, R. Renner, and G. Chiribella, Optimal Universal Programming of Unitary Gates, Phys. Rev. Lett. **125**, 210501 (2020).

[31] A. Bisio, G. Chiribella, G. M. D'Ariano, S. Facchini, and P. Perinotti, Optimal quantum learning of a unitary transformation, Phys. Rev. A **81**, 032324 (2010).

[32] M. Sedlák, A. Bisio, and M. Ziman, Optimal Probabilistic Storage and Retrieval of Unitary Channels, Phys. Rev. Lett. **122**, 170502 (2019).

[33] M. T. Quintino, Q. Dong, A. Shimbo, A. Soeda, and M. Murao, Reversing Unknown Quantum Transformations: Universal Quantum Circuit for Inverting General Unitary Operations, Phys. Rev. Lett. **123**, 210502 (2019).

[34] A. Marzuoli and M. Rasetti, Computing spin networks, Ann. Phys. **318**, 345 (2005).

[35] S. P. Jordan, Permutational quantum computing, Quantum Inf. Comput. **10**, 470 (2010).

[36] M. Planat and R. Ul Haq, The magic of universal quantum computing with permutations, Adv. Math. Phys. **2017,** 5287862 (2017).

[37] V. Havlíček and S. Strelchuk, Quantum Schur Sampling Circuits can be Strongly Simulated, Phys. Rev. Lett. **121,** 060505 (2018).

[38] Y. Ouyang, Y. Shen, and L. Chen, Faster quantum computation with permutations and resonant couplings, Linear Algebra Appl. **592,** 270 (2020).

[39] G. Chiribella and D. Ebler, Optimal quantum networks and one-shot entropies, New J. Phys. **18,** 093053 (2016).

[40] Y. Yang, G. Chiribella, and Q. Hu, Units of rotational information, New J. Phys. **19,** 123003 (2017).

[41] M. T. Quintino, Q. Dong, A. Shimbo, A. Soeda, and M. Murao, Probabilistic exact universal quantum circuits for transforming unitary operations, Phys. Rev. A **100,** 062339 (2019).

[42] W. Fulton and J. Harris, *Representation Theory* (Springer Science & Business Media, Berlin, 1991), Vol.129.

[43] M. Hayashi, *Group Representation for Quantum Theory* (Springer, Berlin, 2017).

[44] A. W. Harrow, quant-ph/0512255.

[45] D. Bacon, I. L. Chuang, and A. W. Harrow, Efficient Quantum Circuits for Schur and Clebsch-Gordan Transforms, Phys. Rev. Lett. **97,** 170502 (2006).

[46] H. Krovi, An efficient high dimensional quantum Schur transform, Quantum **3,** 122 (2019).

[47] I. Schur, Ph.D. dissertation, Friedrich-Wilhelms-Universität zu Berlin, 1901.

[48] $\log := \log_2$.

[49] J. Watrous, Zero-knowledge against quantum attacks, SIAM J. Comput. **39,** 25 (2009).

[50] P. Komar, E. M. Kessler, M. Bishof, L. Jiang, A. S. Sørensen, J. Ye, and M. D. Lukin, A quantum network of clocks, Nat. Phys. **10,** 582 (2014).

[51] T. J. Proctor, P. A. Knott, and J. A. Dunningham, Multiparameter Estimation in Networked Quantum Sensors, Phys. Rev. Lett. **120,** 080501 (2018).

[52] Y. Mo and G. Chiribella, Quantum-enhanced learning of rotations about an unknown direction, New J. Phys. **21,** 113003 (2019).

[53] L. Yu, R. B. Griffiths, and S. M. Cohen, Efficient implementation of bipartite nonlocal unitary gates using prior entanglement and classical communication, Phys. Rev. A **81,** 062315 (2010).

[54] L. Yu and K. Nemoto, Implementation of bipartite or remote unitary gates with repeater nodes, Phys. Rev. A **94,** 022320 (2016).

[55] B. Gendra, E. Ronco-Bonvehi, J. Calsamiglia, R. Munoz-Tapia, and E. Bagan, Quantum Metrology Assisted by Abstention, Phys. Rev. Lett. **110,** 100501 (2013).

[56] Y. Yang, G. Chiribella, and M. Hayashi, Communication cost of quantum processes, IEEE J. Sel. Areas Inf. Theory **1,** 387 (2020).