



PAPER • OPEN ACCESS

Constructing a virtual two-qubit gate by sampling single-qubit operations

To cite this article: Kosuke Mitarai and Keisuke Fujii 2021 *New J. Phys.* **23** 023021

View the [article online](#) for updates and enhancements.

You may also like

- [QUBIC I: Overview and science program](#)
The QUBIC collaboration, J.-Ch. Hamilton, L. Mousset et al.
- [Optimal synthesis of the Fredkin gate in a multilevel system](#)
Wen-Qiang Liu and Hai-Rui Wei
- [Perfect Single Qubit Mirroring Effects on Two and Three Maximally Entangled Qubits](#)
M. Ávila



PAPER

Constructing a virtual two-qubit gate by sampling single-qubit operations

OPEN ACCESS

RECEIVED
1 October 2020REVISED
28 November 2020ACCEPTED FOR PUBLICATION
31 December 2020PUBLISHED
12 February 2021

Original content from
this work may be used
under the terms of the
[Creative Commons
Attribution 4.0 licence](#).

Any further distribution
of this work must
maintain attribution to
the author(s) and the
title of the work, journal
citation and DOI.

Kosuke Mitarai^{1,2,3,*} and Keisuke Fujii^{1,2,4}¹ Graduate School of Engineering Science, Osaka University, 1-3 Machikaneyama, Toyonaka, Osaka 560-8531, Japan² Center for Quantum Information and Quantum Biology, Institute for Open and Transdisciplinary Research Initiatives, Osaka University, Osaka 560-8531, Japan³ JST, PRESTO, 4-1-8 Honcho, Kawaguchi, Saitama 332-0012, Japan⁴ Center for Emergent Matter Science, RIKEN, Wako Saitama 351-0198, Japan

* Author to whom any correspondence should be addressed.

E-mail: mitarai@qc.ee.es.osaka-u.ac.jp and fujii@qc.ee.es.osaka-u.ac.jp**Keywords:** quantum computation, NISQ, quantum channels**Abstract**

We show a certain kind of non-local operations can be simulated by sampling a set of local operations with a quasi-probability distribution when the task of a quantum circuit is to evaluate an expectation value of observables. Utilizing the result, we describe a strategy to decompose a two-qubit gate to a sequence of single-qubit operations. Required operations are projective measurement of a qubit in Pauli basis, and $\pi/2$ rotation around x , y , and z axes. The required number of sampling to get an expectation value of a target observable within an error of ϵ is roughly $O(9^k/\epsilon^2)$, where k is the number of ‘cuts’ performed. The proposed technique enables to perform ‘virtual’ gates between a distant pair of qubits, where there is no direct interaction and thus a number of swap gates are inevitable otherwise. It can also be utilized to improve the simulation of a large quantum computer with a small-sized quantum device, which is an idea put forward by Peng *et al* (2019 arXiv:[1904.00102](#)). This work can enhance the connectivity of qubits on near-term, noisy quantum computers.

1. Introduction

Quantum computers have attracted much attention recently, mainly due to the rapid development of actual hardware [1–3]. The quantum computer that is to appear shortly is called noisy intermediate scale quantum devices, or in short, NISQ devices [4]. We expect NISQ devices to have ~ 100 of qubits with non-negligible noise in the near future. Such devices are believed to be not simulatable by classical computers when the control precision of the qubits is sufficiently high [5–8]. In this sense, NISQ devices have computational power that exceeds classical computers. Many researchers are actively developing ways to exploit their power for practical applications [9–15]. However, we still suffer from the limited number of qubits available on actual devices and the limited depth of circuits that can be run while maintaining the resultant quantum state meaningful.

If techniques to decompose a quantum circuit to smaller ones are developed, they can extend the applicability of such devices. Smaller quantum circuits may refer to ones with the smaller number of qubits or gates. Peng *et al* recently proposed a clustering approach based on a tensor network representation of a quantum circuit [16], which greatly progressed the technical development. They showed that we can ‘cut’ an identity gate, by sampling measure-and-prepare channels on a qubit according to a certain quasi-probability distribution. In reference [17], we proposed methods to construct quantum circuits equivalent to the Hadamard test, which successfully reduces the depth of certain quantum circuits. These techniques share a same idea in that they reconstruct a result of a coherent quantum operation from certain incoherent operations by combining the results obtained from them.

An approach which has the same flavor as the above have been utilized in the context of memory-efficient classical simulation of quantum circuits. Since the direct simulation of a quantum circuit

with over 50 qubits breaks down due to the need of storing 2^{50} complex numbers in memory, the classical simulator must decompose the given quantum circuit to smaller ones, especially in the number of qubits. References [18, 19] have provided one way for such decomposition, which ‘cuts’ controlled-Z gates by separately simulating two cases where the control qubit is $|0\rangle$ or $|1\rangle$ and then combining them, and they performed classical simulation of over 50-qubit quantum circuits. A similar technique has been utilized by Bravyi *et al* in reference [20] to remove a relatively small number of qubits from a large quantum circuit by replacing the qubits with a classical simulator. Their approach can be viewed as ‘space-like’ cut rather than the ‘time-like’ cut proposed by Peng *et al* [16]. However, their techniques are intended to run on a classical computer and cannot be utilized for simulating a large quantum circuit with a small quantum computer.

In this work, we present a technique to perform ‘space-like’ cut on a quantum computer. More specifically, we present a way to decompose a controlled gate into a sequence of single-qubit operations which consists of projective measurements of Pauli X, Y, and Z operators, and single-qubit rotations around x , y , and z -axes. We note that our method does not generate any entanglement between the qubits as it is impossible to do so with such single-qubit operations. Our method only ‘simulates’ effects of entanglement using classical post-processing and sampling. More concretely, although entangling gates cannot be performed with local operations and classical communications in single-shot experiments as widely known [21], we show that it is possible to perform a computational task of evaluating expectation values of the output of entangling circuits by sampling certain sets of gates and applying classical post-processing. The overhead required for our proposed technique, which scales exponentially to the number of decomposition performed, gives a characterization of the entangling gates from a computational viewpoint, which is different from the existing theories of entanglement quantification in e.g. [22].

The method proposed here can be considered as a generalization of our previous work [17] and a variant of the quantum circuit decomposition presented in reference [16]. It can also be viewed as a fully quantum version of the technique utilized in efficient classical simulation schemes [18–20]. In some cases, our method provides a better scaling against reference [16] when simulating a large quantum circuit with smaller ones. The proposed technique is also useful when we want to apply two-qubit gates between a distant pair of qubits, which otherwise would require many swap operations to perform. This work extends the applicability of NISQ devices whose circuit depth and connectivity are limited.

2. Gate decomposition

2.1. Tensor network representation of quantum circuits

Quantum computation is completely specified with a quantum circuit, U , an initial state with its density matrix representation, ρ , and an observable, O , measured at the output. Given U , ρ , and O , any quantum computation can be represented by a tensor network [23–25]. We define the tensor representation of U , ρ , and O in the following manner.

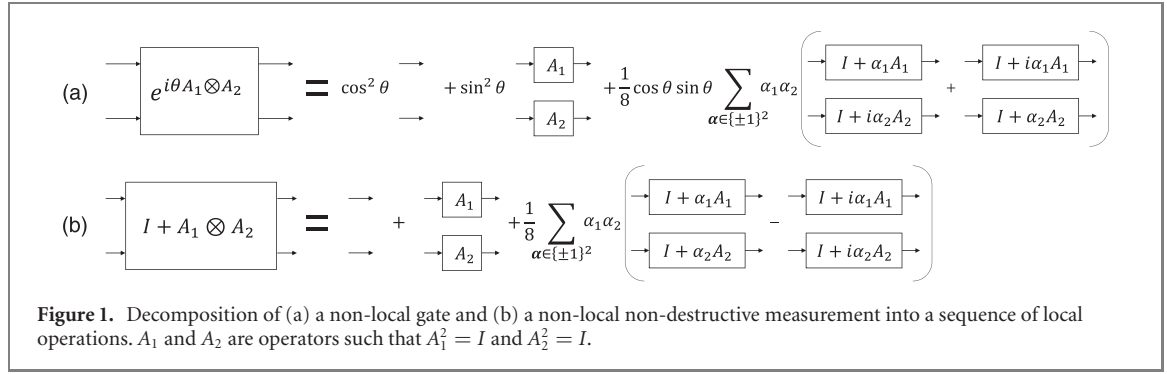
Suppose that our quantum computer has n qubits. We define a complete set of basis in the space of 2×2 complex matrix and its dual as $\{|e_i\rangle\}_{i=1}^4$ and $\{\langle e_i|\}_{i=1}^4$ respectively, and assume orthonormality under the trace inner product; $\langle e_i|e_j\rangle = \delta_{ij}$. We use the trace inner product, that is, for matrices A and B , $\langle A|B\rangle = \text{Tr}(A^\dagger B)$. A density matrix ρ can be decomposed into the sum of $|e_{j_1}\rangle \otimes |e_{j_2}\rangle \otimes \cdots \otimes |e_{j_n}\rangle = |e_{j_1} e_{j_2} \cdots e_{j_n}\rangle$ as

$$|\rho\rangle\rangle = \sum_{j_1, \dots, j_n} \rho_j |e_{j_1} e_{j_2} \cdots e_{j_n}\rangle, \quad (1)$$

where $\mathbf{j} = (j_1, j_2, \dots, j_n)$. We refer to the elements $\rho_j = \langle e_{j_1} e_{j_2} \cdots e_{j_n} | \rho \rangle$ as the tensor representation of ρ . An observable O can also be decomposed into the same form. Note that we can naturally assume tensor representations of observables and density matrices consist of real numbers because they are always Hermitian and we can choose the basis $\{|e_i\rangle\}_{i=1}^4$ as Hermitian, e.g. we can use the Pauli matrices $\{I, X, Y, Z\}$ as the basis. Therefore, we assume ρ_j and O_i are real henceforth. The quantum circuit, U , transforms ρ into $U\rho U^\dagger$. We define a corresponding superoperator $\mathcal{S}(U)$ whose action is defined by $\mathcal{S}(U)\rho = U\rho U^\dagger$. Superoperator can be decomposed as,

$$\mathcal{S}(U) = \sum_{j_1, \dots, j_n} \sum_{k_1, \dots, k_n} \mathcal{S}(U)_{j,k} |e_{j_1} \cdots e_{j_n}\rangle \langle e_{k_1} \cdots e_{k_n}|. \quad (2)$$

Note that this decomposition is not limited to superoperators of unitary matrices, but also is applicable for any linear operator that acts on a density matrix. We call $\mathcal{S}(U)_{j,k} = \langle e_{j_1} \cdots e_{j_n} | \mathcal{S}(U) | e_{k_1} \cdots e_{k_n} \rangle$ tensor representation of $\mathcal{S}(U)$. When we use the Pauli operators as basis set, $\mathcal{S}(U)_{j,k}$ is referred as Pauli transfer matrix.



Quantum computation ends with measuring the observable O . This output can be written down as,

$$\langle\langle O | \mathcal{S}(U) | \rho \rangle\rangle = \text{Tr}(OU\rho U^\dagger) \quad (3)$$

$$= \sum_{j_1, \dots, j_n} \sum_{k_1, \dots, k_n} O_j \mathcal{S}(U)_{j,k} \rho_k, \quad (4)$$

In many cases, U is a product of elementary gates $\{U_i\}_{i=1}^L$, that is, $U = U_L \dots U_1$. The tensor representation of the overall gate, $\mathcal{S}(U)$, is also a product of $\mathcal{S}(U_i)$; $\mathcal{S}(U) = \mathcal{S}(U_L) \dots \mathcal{S}(U_1)$. An important note is that as long as the tensor representation of each element is unchanged, the result of the overall computation is also unchanged. If $\mathcal{S}(U)$ can be represented by a sum of some simple operations as $\mathcal{S}(U) = \sum_i c_i \mathcal{S}(V_i)$ with coefficients $\{c_i\}$, the expectation value of an observable O can be computed with the following equality,

$$\langle\langle O | \mathcal{S}(U) | \rho \rangle\rangle = \sum_i c_i \langle\langle O | \mathcal{S}(V_i) | \rho \rangle\rangle. \quad (5)$$

Note that c_i can, in general, depend on the state $|\rho\rangle$. We use this scheme to perform the ‘decomposition’ of a circuit in this work.

It is noteworthy that as we perform decompositions of a superoperator rather than an operator such as U itself, the method becomes friendly for a realistic quantum device. A direct decomposition of U into some simple operators $\{V_i\}$, i.e. $U = \sum_i c_i V_i$, can also be utilized for the same task; however, as expectation values are calculated as $\langle 0 | U^\dagger O U | 0 \rangle$ where $|0\rangle$ is an initial state, this approach requires us to evaluate $\sum_{i,j} c_i c_j^* \langle 0 | V_j^\dagger O V_i | 0 \rangle$ which are rather hard for the NISQ devices. This fact demonstrates the advantage of using the above formalism. The tensor network representation of the superoperator formalism allows us to graphically understand the decompositions.

2.2. Virtual two-qubit gate

We can show the following, which can then be utilized to decompose any two-qubit gate into a sequence of single-qubit operations.

Lemma 1. For operators A_1 and A_2 such that $A_1^2 = I$ and $A_2^2 = I$,

$$\begin{aligned} \mathcal{S}(e^{i\theta A_1 \otimes A_2}) &= \cos^2 \theta \mathcal{S}(I \otimes I) + \sin^2 \theta \mathcal{S}(A_1 \otimes A_2) \\ &+ \frac{1}{8} \cos \theta \sin \theta \sum_{(\alpha_1, \alpha_2) \in \{\pm 1\}^2} \alpha_1 \alpha_2 [\mathcal{S}((I + \alpha_1 A_1) \otimes (I + i \alpha_2 A_2)) \\ &+ \mathcal{S}((I + i \alpha_1 A_1) \otimes (I + \alpha_2 A_2))] \end{aligned} \quad (6)$$

To prove this, we can directly check the tensor representation of both hand side is equivalent. For detailed calculation, see appendix A1. This theorem is schematically depicted in figure 1(a). Notice that the operation that is proportional to $I \pm A$ and $I \pm iA$ for $A \in \{X, Y, Z\}$ can respectively be performed by a projective measurement and a single-qubit rotation.

The correspondence with a single-qubit rotation is clear from the formula, $e^{\pm i\pi A/4} = \frac{1}{\sqrt{2}}(I \pm iA)$, which is the rotation of angle $\pi/2$ around the A axis. Let \mathcal{M}_A be the projective measurement on the A basis ($A \in \{X, Y, Z\}$), that is, \mathcal{M}_A acts on a density matrix ρ as,

$$\mathcal{M}_A \rho = \frac{1}{\text{Tr}(\rho \frac{I + \alpha A}{2})} \left(\frac{I + \alpha A}{2} \right) \rho \left(\frac{I + \alpha A}{2} \right), \quad (7)$$

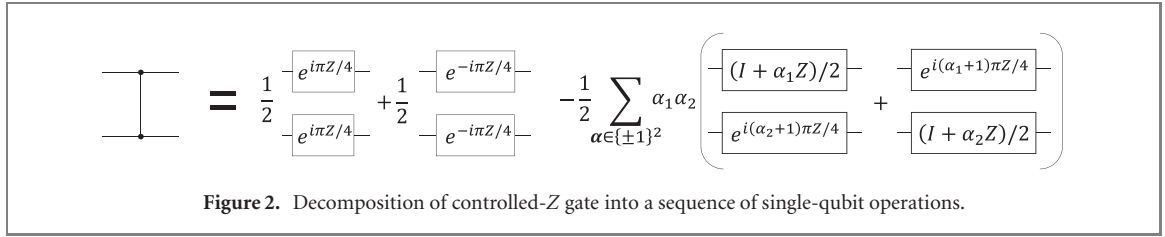


Figure 2. Decomposition of controlled-Z gate into a sequence of single-qubit operations.

depending on the result of the measurement $\alpha \in \{1, -1\}$. This is equivalent to $S(I \pm A)$ up to the factor of $4 \text{Tr}(\rho \frac{I+\alpha A}{2})$, that is,

$$S(I + \alpha A) = 4 \text{Tr} \left(\rho \frac{I + \alpha A}{2} \right) \mathcal{M}_{A,\alpha}, \quad (8)$$

where $\mathcal{M}_{A,\alpha}$ is a measurement operation postselected with the measurement outcome α . $\text{Tr}(\rho \frac{I+\alpha A}{2})$ is the probability of getting the result α by measuring ρ on the A basis. Lemma 1 with this fact implies that the gate $e^{i\theta A_1 \otimes A_2}$ can be decomposed, in a sense of equation (5), into a sum of $I \otimes I$, $A_1 \otimes A_2$, $\mathcal{M}_{A_1} \otimes e^{\pm i\pi A_2/4}$, and $e^{\pm i\pi A_1/4} \otimes \mathcal{M}_{A_2}$, which can be stated as lemma below. Notably, this technique can be applied for any θ , which enables us to perform continuous two-qubit gates.

Lemma 2. *A quantum gate $e^{i\theta A_1 \otimes A_2}$ with operators A_1 and A_2 such that $A_1^2 = I$ and $A_2^2 = I$ can be decomposed into six single-qubit operations. For any quantum state $|\rho\rangle$, to achieve the error ϵ of the decomposition with respect to the trace distance with probability at least $1 - \delta$, the required number of circuit runs is $O(\log(1/\delta)/\epsilon^2)$.*

The detailed proof is given in appendix B. Intuitively, since the error comes from the probabilistic part of the decomposition, that is the renormalization factor in equation (8) $\text{Tr}(\rho \frac{I+\alpha A}{2})$, if we want to estimate $\text{Tr}(\rho \frac{I+\alpha A}{2})$ within error ϵ , $O(1/\epsilon^2)$ repetition would suffice.

Let us finally mention the case of the controlled-Z gate, which we denote by CZ. CZ can be decomposed into

$$\text{CZ} = e^{i\pi I \otimes Z/4} e^{i\pi Z \otimes I/4} e^{-i\pi Z \otimes Z/4}, \quad (9)$$

ignoring the global phase. This means we can decompose a CZ gate using lemma 2. The decomposition is shown in figure 2. Similar decompositions can be performed on some basic two-qubit gates such as CNOT. Endo *et al* [26] also provides such decomposition (reference [26], appendix B). However, our protocol above is slightly advantageous in that the number of single-qubit operations required is six compared to theirs which requires nine of them.

2.3. Virtual non-destructive measurement of two-qubit operators

In the previous subsection, we showed that any two-qubit rotation can be decomposed into a sum of single-qubit operations. Here, we extend the strategy to construct virtual non-destructive measurement of two-qubit operators. Similar to the previous section, we can show the following. This theorem is schematically shown in figure 1(b).

Lemma 3. *For operators A_1 and A_2 such that $A_1^2 = I$ and $A_2^2 = I$,*

$$S(I + A_1 \otimes A_2) = S(I \otimes I) + S(A_1 \otimes A_2) + \frac{1}{8} \sum_{(\alpha_1, \alpha_2) \in \{\pm 1\}^2} \alpha_1 \alpha_2 [S((I + \alpha_1 A_1) \otimes (I + \alpha_2 A_2)) - S((I + i\alpha_1 A_1) \otimes (I + i\alpha_2 A_2))]. \quad (10)$$

This can also be shown by the direct calculation of both hand side. See appendix A2 for detailed calculation.

The above lemma can be utilized to show the following.

Lemma 4. *A non-local projection $\frac{I+A_1 \otimes A_2}{2}$ with operators A_1 and A_2 such that $A_1^2 = 1$ and $A_2^2 = 2$ can be decomposed into six single-qubit operations. For any quantum state $|\rho\rangle$, to achieve the error ϵ of the decomposition with respect to the trace distance with probability at least $1 - \delta$, the required number of circuit runs is $O(\log(1/\delta)/\epsilon^2)$.*

This can be shown with exactly the same approach taken to prove lemma 2, which is provided in appendix B.

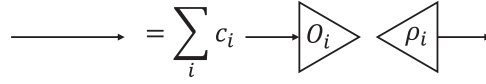


Figure 3. Time-like cut employed in reference [16].

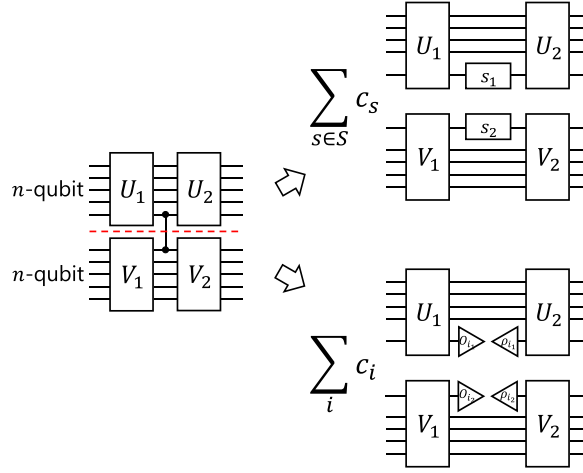


Figure 4. Two decomposition approach compared in main text. The top-right approach is the presented, and the bottom-right approach is of reference [16].

3. Application

3.1. Simulation of large quantum circuits

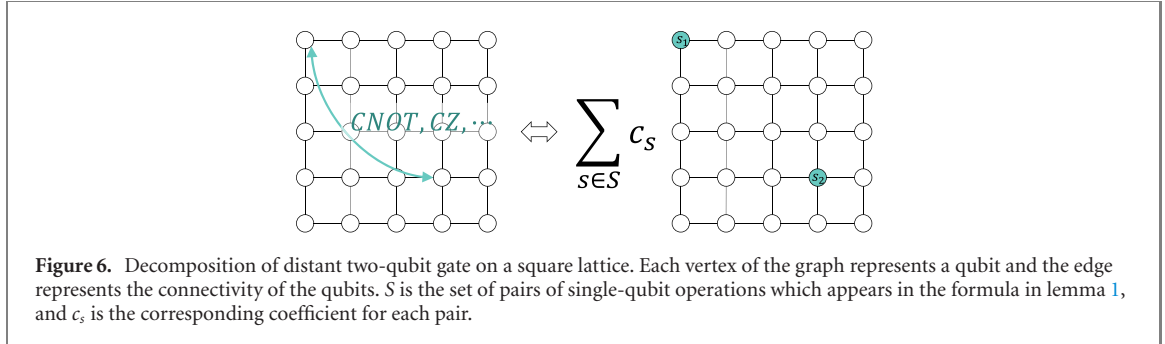
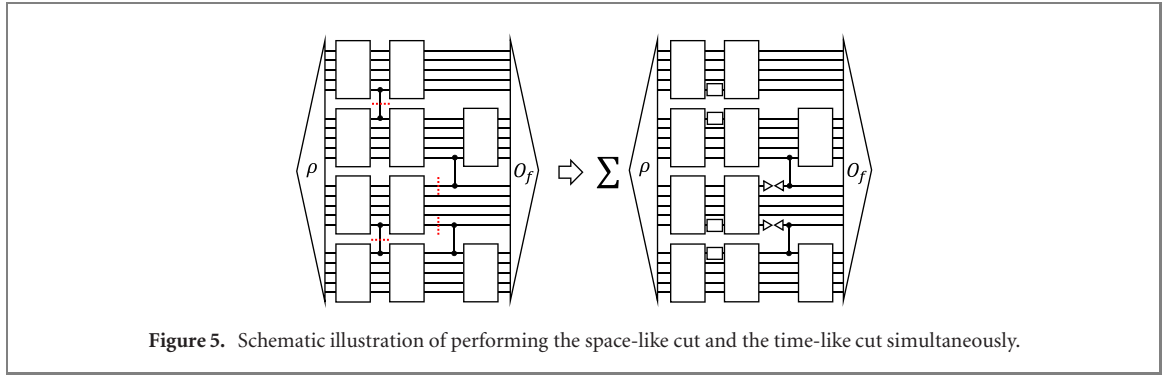
The idea of simulating a large quantum circuit by a small quantum computer has been put forward in reference [16]. Peng *et al* utilized the equivalence shown in figure 3. In the figure,

$$\begin{aligned}
 O_1 &= I, \rho_1 = |0\rangle\langle 0|, & c_1 &= +1/2, \\
 O_2 &= I, \rho_2 = |1\rangle\langle 1|, & c_2 &= +1/2, \\
 O_3 &= X, \rho_3 = |+\rangle\langle +|, & c_3 &= +1/2, \\
 O_4 &= X, \rho_4 = |-\rangle\langle -|, & c_4 &= -1/2, \\
 O_5 &= Y, \rho_5 = |+i\rangle\langle +i|, & c_5 &= +1/2, \\
 O_6 &= Y, \rho_6 = |-i\rangle\langle -i|, & c_6 &= -1/2, \\
 O_7 &= Z, \rho_7 = |0\rangle\langle 0|, & c_7 &= +1/2, \\
 O_8 &= Z, \rho_8 = |1\rangle\langle 1|, & c_8 &= -1/2,
 \end{aligned} \tag{11}$$

where $|\pm\rangle = (|0\rangle \pm |1\rangle)/\sqrt{2}$ and $|\pm i\rangle = (|0\rangle \pm i|1\rangle)/\sqrt{2}$. The symbols \triangleright and \triangleleft denotes the measurement of a certain observable and the preparation of a certain state, respectively. Contrasting this technique and ours, we refer to the former and the latter as ‘time-like’ and ‘space-like’ cut, respectively. More concretely, a *time-like cut* of a quantum channel can be defined as a decomposition of the channel in the sense of equation (5) using measure-and-prepare channels only. In contrast, a *space-like cut* of a non-local quantum channel is a decomposition of the channel using local quantum channels only.

The decomposition presented in the previous section can also be used in this direction. Let us compare the scaling of cost of our decomposition scheme and that of Peng *et al* by a simple example. We consider the case where we have an n -qubit quantum computer to simulate a $2n$ -qubit quantum circuit of figure 4, which has only one CZ gate between n -qubit ‘cluster’. The task is to estimate the expectation value of a final observable O_f by measuring it in the computational basis. To simplify the discussion, we assume O_f is a string of Pauli Z ’s.

Let v be a desired variance of the estimation of the expectation value of O_f . We can show a naive algorithm, which runs the equal number of circuits for each terms appearing in the decomposition, to perform the decomposition with time-like cuts, in the worst case, requires $2048/v$ runs of n -qubit circuit,



while the space-like cut approach takes $\frac{15}{2v}$ runs. The analysis of this simple example is given in appendix D. Although the analysis given here is based on a naive algorithm and there are possibilities to improve it, this analysis somewhat shows the enhancement provided by our space-like cut protocol.

General case.

We can consider a general case where we perform the time-like and space-like cuts simultaneously to make a given m -qubit quantum circuit runnable on an n -qubit quantum computer. Let the number of time-like and space-like cuts be M_t and M_s , respectively. See figure 5 for a schematic illustration. For space-like cuts, we assume they are performed only on CZ gates. The input state ρ is initialized in $|0\rangle\langle 0|^{\otimes m}$ and O_f is an output (diagonal) observable calculated from some output function $f: \{0, 1\}^m \rightarrow [-1, 1]$. Our task here is to estimate the expectation $\mathbb{E}[f(y)]$ for a random bitstring $y \in \{0, 1\}^m$ sampled from the original circuit. This model is adopted from reference [16] which originates in reference [20]. With this definition, we can get the following.

Theorem 5. *The number of n -qubit circuit runs required to estimate $\mathbb{E}[f(y)]$ within accuracy ϵ with some high probability $1 - \delta$ is $O\left(\frac{9^{M_s} 16^{M_t}}{\epsilon^2} \log\left(\frac{1}{2\delta}\right)\right)$.*

This implies that the decomposition of the circuit should be performed to minimize $9^{M_s} 16^{M_t}$. A detailed proof is given in appendix E, however, the above can roughly be explained as follows. At each space-like cut, we get six different sets of single-qubit operations, so M_s cuts induce 6^{M_s} terms. Likewise, M_t time-like cuts induce 8^{M_t} terms, which makes the total number of circuits in decomposition $6^{M_s} 8^{M_t}$. With this decomposition, we can take a Monte-Carlo approach to estimate the sum, that is, we randomly choose circuits to run and average them. Hoeffding's inequality can be used to bound the error of such protocol, which states that if a magnitude of a random variable is always bounded by some constant a , then $O(a^2/\epsilon^2)$ samples would suffice to obtain an accuracy of ϵ . In this case, we are to estimate $\mathbb{E}[f(y)] = \sum_{i=1}^{6^{M_s} 8^{M_t}} c_i \langle O_f | \mathcal{S}(V_i) | \rho \rangle$ with i randomly drawn from $\{1, \dots, 6^{M_s} 8^{M_t}\}$ and $|c_i| = 1/2^{M_s+M_t}$, that is, $\mathbb{E}[f(y)]$ is estimated by $\mathbb{E}_i[6^{M_s} 8^{M_t} c_i \langle O_f | \mathcal{S}(V_i) | \rho \rangle]$. The magnitude of random variable $6^{M_s} 8^{M_t} c_i \langle O_f | \mathcal{S}(V_i) | \rho \rangle$ is roughly $3^{M_s} 4^{M_t}$, thus we can apply the Hoeffding bound to get the result.

3.2. Distant two-qubit gates

The theorem introduced above can be utilized to 'virtually' perform a two-qubit gate between qubits at distance. Figure 6 shows an example of such a virtual two-qubit gate. Notice that this protocol works

irrespective of the distance between the qubits. Many swap gates are otherwise necessary for performing such gates, which makes them impractical on NISQ devices due to the non-negligible amount of decoherence and gate error of such devices.

This protocol might be useful for the variational algorithms such as the variational quantum eigensolver (VQE) [9] and the quantum approximate optimization algorithms (QAOA) [12]. Here, we describe an example in the QAOA. In the QAOA, we seek to find a ground state of a Hamiltonian H on n -qubit which is a sum of Pauli Z 's and its products. For example, a Hamiltonian may have the form of,

$$H = \sum_{ij} J_{ij} Z_i Z_j. \quad (12)$$

The QAOA tries to solve the problem by converting it to an optimization problem of a continuous variable β and γ . The optimization of β and γ are performed so as to minimize the function,

$$\langle H(\beta, \gamma) \rangle = \langle + |^{\otimes n} U^\dagger(\beta, \gamma) H U(\beta, \gamma) | + \rangle^{\otimes n}, \quad (13)$$

where,

$$U(\beta, \gamma) = e^{i\beta_p \sum_i X_i} e^{i\gamma_p H} \dots e^{i\gamma_2 H} e^{i\beta_1 \sum_i X_i} e^{i\gamma_1 H}. \quad (14)$$

This algorithm has been experimentally demonstrated [13] with the connectivity of the target Hamiltonian being equivalent to the connectivity of the actual device.

The equivalence of the connectivity is almost necessary from the requirement to perform $e^{i\gamma H}$. This requirement can somewhat be relaxed by our protocol which enables qubits to virtually interact irrespective of the distance between them. Let us now assume that an available device has a square-lattice connectivity of figure 6, and a Hamiltonian of the QAOA which we aim to solve has an interaction between one pair of qubits that is not included in the hardware connectivity graph. In this case, to execute the QAOA circuit (equation (14)), we can use our space-like technique p times to virtually apply the unitary. The scaling of the cost can be bounded by setting $M_t = 0$ and $M_s = p$ in theorem 5 which gives us a scaling of $9^p \epsilon^{-2} \log[1/(2\delta)]$. The time-like cut approach of Peng *et al* [16] can also be utilized in this direction. However, as this approach would require 4 cuts per gate, the cost scaling is bounded by $16^{4p} \epsilon^{-2} \log[1/(2\delta)]$ by setting $M_t = 4p$ and $M_s = 0$ in theorem 5. This demonstrates an advantage, albeit in this special settings, of our technique over the previous result.

In the context of the VQE, which is also an algorithm to find a ground state of a Hamiltonian but mainly targets a concrete physical system such as molecules, it has been proposed to use the same kind of quantum circuits as the QAOA [27, 28]. Our result may also be applicable in constructing such circuits.

4. Discussion and conclusion

We described a technique to decompose a non-local operations into a sequence of local operations. As the single-qubit operations are generally more accurate on NISQ devices, the proposed technique can be used to enhance their capability. We believe intrinsic noise on single-qubit operations can be compensated by recent sophisticated error mitigation techniques [26]. In particular, our technique of the space-like cut of two-qubit gates can improve the simulation of a large quantum circuit with a small quantum computer in some cases. It would be interesting to investigate the best strategy to perform 'cuts' to reduce the number of qubits compatible with an available device. Also, the algorithm we have given to bound the cost scaling is rather straight forward and we believe it can be improved with a more sophisticated strategy.

The proposed algorithm can also be compared to the classical simulation strategy that splits a large circuit by decomposing two-qubit gates. For example, a controlled-NOT gate can be splitted using a tensor network based technique [29]. However, such techniques generally does not focus on decompositions of $\mathcal{S}(U)$ considered in this work but rather the two-qubit unitary U itself, which takes makes them difficult to be used on NISQ devices as equation (5) cannot be utilized anymore.

Our technique can induce an entanglement-like effect without performing any two-qubit gate with the cost mentioned in lemmas 2 and 4. This connects this work to areas like quantum communication. This 'virtual' entanglement creation could be done with the time-like cut proposed by Peng *et al*, but our work lowered the cost to perform the task. It is interesting to know whether ours is the optimal protocol or there is a more efficient way.

To summarize, our technique allows qubits to virtually interact irrespective of physical distances between them. The result is useful for applying a two-qubit gate to a distant pair of qubits. In particular,

when applied to the NISQ devices, this may be employed to enhance the power of them. Future direction can be to explore if we can lower the resource to perform such virtual operations.

Acknowledgments

KM thanks the METI and IPA for their support through the MITOU Target program. KM is also supported by JSPS KAKENHI No. 19J10978 and No. 20K22330, and JST PRESTO JPMJPR2019. KF is supported by KAKENHI No. 16H02211, JST PRESTO JPMJPR1668, JST ERATO JPMJER1601, and JST CREST JPMJCR1673. The authors thank Suguru Endo for fruitful discussions and letting us become aware of reference [26]. This work is supported by MEXT Quantum Leap Flagship Program (MEXT Q-LEAP) Grant Nos. JPMXS0118067394 and JPMXS0120319794.

Appendix A. Proof of lemmas 1 and 3

A tensor representation of $\mathcal{S}((I + \alpha_1 A_1) \otimes (I + \alpha_2 A_2))$ on a set of basis $\{|e_i e_j\rangle\}_{i,j=1}^4$ is as follows.

$$\begin{aligned}
 & \langle\langle e_i e_j | \mathcal{S}(I + \alpha_1 A_1) \otimes (I + \alpha_2 A_2) | e_k e_l \rangle\rangle \\
 &= \text{Tr} (e_i \otimes e_j (I + \alpha_1 A_1) \otimes (I + \alpha_2 A_2) e_k \otimes e_l (I + \alpha_1^* A_1) \otimes (I + \alpha_2^* A_2)) \\
 &= \text{Tr} (e_i \otimes e_j e_k \otimes e_l) + \alpha_1 \text{Tr} (e_i \otimes e_j (A_1 \otimes I) e_k \otimes e_l) + \alpha_1^* \text{Tr} (e_k \otimes e_l (A_1 \otimes I) e_i \otimes e_j) \\
 &\quad + \alpha_2 \text{Tr} (e_i \otimes e_j (I \otimes A_2) e_k \otimes e_l) + \alpha_2^* \text{Tr} (e_k \otimes e_l (I \otimes A_2) e_i \otimes e_j) \\
 &\quad + \alpha_1 \alpha_2 \text{Tr} (e_i \otimes e_j (A_1 \otimes A_2) e_k \otimes e_l) + \alpha_1^* \alpha_2^* \text{Tr} (e_k \otimes e_l (A_1 \otimes A_2) e_i \otimes e_j) \\
 &\quad + \alpha_1 \alpha_2^* \text{Tr} (e_i \otimes e_j (A_1 \otimes I) e_k \otimes e_l (I \otimes A_2)) + \alpha_1^* \alpha_2 \text{Tr} (e_i \otimes e_j (I \otimes A_2) e_k \otimes e_l (A_1 \otimes I)) \\
 &\quad + \alpha_1 \text{Tr} (e_i \otimes e_j (A_1 \otimes A_2) e_k \otimes e_l (I \otimes A_2)) + \alpha_1^* \text{Tr} (e_i \otimes e_j (I \otimes A_2) e_k \otimes e_l (A_1 \otimes A_2)) \\
 &\quad + \alpha_2 \text{Tr} (e_i \otimes e_j (A_1 \otimes A_2) e_k \otimes e_l (A_1 \otimes I)) + \alpha_2^* \text{Tr} (e_i \otimes e_j (A_1 \otimes I) e_k \otimes e_l (A_1 \otimes A_2)) \\
 &\quad + \text{Tr} (e_i \otimes e_j (A_1 \otimes A_2) e_k \otimes e_l (A_1 \otimes I)) + \text{Tr} (e_i \otimes e_j (A_1 \otimes I) e_k \otimes e_l (A_1 \otimes A_2)) \\
 &\quad + \text{Tr} (e_i \otimes e_j (A_1 \otimes A_2) e_k \otimes e_l (A_1 \otimes A_2)). \tag{A1}
 \end{aligned}$$

Let,

$$\{\alpha_1, \alpha_2\}_{ij,kl} := \mathcal{S}((I + \alpha_1 A_1) \otimes (I + \alpha_2 A_2))_{ij,kl}, \tag{A2}$$

$$\begin{pmatrix} a_{1,ijkl} \\ a_{2,ijkl} \\ a_{3,ijkl} \\ a_{4,ijkl} \\ a_{5,ijkl} \\ a_{6,ijkl} \\ a_{7,ijkl} \\ a_{8,ijkl} \\ a_{9,ijkl} \\ a_{10,ijkl} \\ a_{11,ijkl} \\ a_{12,ijkl} \\ a_{13,ijkl} \\ a_{14,ijkl} \\ a_{15,ijkl} \\ a_{16,ijkl} \end{pmatrix} := \begin{pmatrix} \text{Tr} (e_i \otimes e_j e_k \otimes e_l) \\ \text{Tr} (e_i \otimes e_j (A_1 \otimes I) e_k \otimes e_l) \\ \text{Tr} (e_k \otimes e_l (A_1 \otimes I) e_i \otimes e_j) \\ \text{Tr} (e_i \otimes e_j (I \otimes A_2) e_k \otimes e_l) \\ \text{Tr} (e_k \otimes e_l (I \otimes A_2) e_i \otimes e_j) \\ \text{Tr} (e_i \otimes e_j (A_1 \otimes A_2) e_k \otimes e_l) \\ \text{Tr} (e_k \otimes e_l (A_1 \otimes A_2) e_i \otimes e_j) \\ (e_i \otimes e_j (A_1 \otimes I) e_k \otimes e_l (I \otimes A_2)) \\ (e_i \otimes e_j (I \otimes A_2) e_k \otimes e_l (A_1 \otimes I)) \\ (e_i \otimes e_j (A_1 \otimes A_2) e_k \otimes e_l (I \otimes A_2)) \\ (e_i \otimes e_j (I \otimes A_2) e_k \otimes e_l (A_1 \otimes A_2)) \\ (e_i \otimes e_j (A_1 \otimes A_2) e_k \otimes e_l (A_1 \otimes I)) \\ (e_i \otimes e_j (A_1 \otimes I) e_k \otimes e_l (A_1 \otimes A_2)) \\ (e_i \otimes e_j (A_1 \otimes I) e_k \otimes e_l (A_1 \otimes I)) \\ (e_i \otimes e_j (I \otimes A_2) e_k \otimes e_l (I \otimes A_2)) \\ (e_i \otimes e_j (A_1 \otimes A_2) e_k \otimes e_l (A_1 \otimes A_2)) \end{pmatrix}. \tag{A3}$$

The relation can be summarized in matrix form,

$$\begin{pmatrix} \{+1, +1\}_{ij,kl} \\ \{+1, +i\}_{ij,kl} \\ \{+1, -1\}_{ij,kl} \\ \{+1, -i\}_{ij,kl} \\ \{+i, +1\}_{ij,kl} \\ \{+i, +i\}_{ij,kl} \\ \{+i, -1\}_{ij,kl} \\ \{+i, -i\}_{ij,kl} \\ \{-1, +1\}_{ij,kl} \\ \{-1, +i\}_{ij,kl} \\ \{-1, -1\}_{ij,kl} \\ \{-1, -i\}_{ij,kl} \\ \{-i, +1\}_{ij,kl} \\ \{-i, +i\}_{ij,kl} \\ \{-i, -1\}_{ij,kl} \\ \{-i, -i\}_{ij,kl} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & i & -i & i & -i & -i & i & 1 & 1 & i & -i & 1 & 1 & 1 \\ 1 & 1 & 1 & -1 & -1 & -1 & -1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 & 1 \\ 1 & 1 & 1 & -i & i & -i & i & i & -i & 1 & 1 & -i & i & 1 & 1 & 1 \\ 1 & i & -i & 1 & 1 & i & -i & i & -i & i & -i & 1 & 1 & 1 & 1 & 1 \\ 1 & i & -i & i & -i & -1 & -1 & 1 & 1 & i & -i & i & -i & 1 & 1 & 1 \\ 1 & i & -i & -1 & -1 & -i & i & -i & i & i & -i & -1 & -1 & 1 & 1 & 1 \\ 1 & i & -i & -i & i & 1 & 1 & -1 & -1 & i & -i & -i & i & 1 & 1 & 1 \\ 1 & -1 & -1 & 1 & 1 & -1 & -1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 & 1 \\ 1 & -1 & -1 & i & -i & -i & i & i & -i & -1 & -1 & i & -i & 1 & 1 & 1 \\ 1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 \\ 1 & -1 & -1 & -i & i & i & -i & -i & i & -1 & -1 & -i & i & 1 & 1 & 1 \\ 1 & -i & i & 1 & 1 & -i & i & -i & i & -i & i & 1 & 1 & 1 & 1 & 1 \\ 1 & -i & i & i & -i & 1 & 1 & -1 & -1 & -i & i & i & -i & 1 & 1 & 1 \\ 1 & -i & i & -1 & -1 & i & -i & i & -i & -i & i & -1 & -1 & 1 & 1 & 1 \\ 1 & -i & i & -i & i & -1 & -1 & 1 & 1 & -i & i & -i & i & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \\ a_5 \\ a_6 \\ a_7 \\ a_8 \\ a_9 \\ a_{10} \\ a_{11} \\ a_{12} \\ a_{13} \\ a_{14} \\ a_{15} \\ a_{16} \end{pmatrix}. \quad (\text{A4})$$

A.1. Proof of lemma 1

Tensor representation of $\mathcal{S}(e^{i\theta A_1 \otimes A_2})$ is,

$$\begin{aligned} \langle e_i e_j | \mathcal{S}(e^{i\theta A_1 \otimes A_2}) | e_k e_l \rangle &= \langle e_i e_j | \mathcal{S}(\cos \theta + i \sin \theta A_1 \otimes A_2) | e_k e_l \rangle \\ &= \text{Tr} (e_i \otimes e_j (\cos \theta I + i \sin \theta A_1 \otimes A_2) e_k \otimes e_l (\cos \theta I - i \sin \theta A_1 \otimes A_2)) \\ &= \cos^2 \theta a_{1,ijkl} + i \sin \theta \cos \theta (a_{6,ijkl} - a_{7,ijkl}) + \sin^2 \theta a_{16,ijkl}. \end{aligned} \quad (\text{A5})$$

Observe that,

$$\begin{aligned} &\{+1, +i\} - \{+1, -i\} \\ &= 2i(a_4 - a_5) + 2i(a_6 - a_7) - 2i(a_8 - a_9) + 2i(a_{12} - a_{13}), \\ &\{-1, +i\} - \{-1, -i\} \\ &= 2i(a_4 - a_5) - 2i(a_6 - a_7) + 2i(a_8 - a_9) + 2i(a_{12} - a_{13}), \\ &\{+i, +1\} - \{-i, +1\} \\ &= 2i(a_2 - a_3) + 2i(a_6 - a_7) + 2i(a_8 - a_9) + 2i(a_{12} - a_{13}), \\ &\{+i, -1\} - \{-i, -1\} \\ &= 2i(a_2 - a_3) - 2i(a_6 - a_7) - 2i(a_8 - a_9) + 2i(a_{12} - a_{13}), \end{aligned} \quad (\text{A6})$$

where we abbreviated the subscripts $ijkl$. We can solve the above for $i(a_6 - a_7)$, and obtain

$$\begin{aligned} 8i(a_6 - a_7) &= \{+1, +i\} - \{+1, -i\} - \{-1, +i\} + \{-1, -i\} \\ &\quad + \{+i, +1\} - \{-i, +1\} - \{+i, -1\} + \{-i, -1\} \end{aligned} \quad (\text{A7})$$

$$= \sum_{\alpha \in \{\pm 1\}^2} \alpha_1 \alpha_2 [\mathcal{S}((I + \alpha_1 A_1) \otimes (I + i\alpha_2 A_2)) + \mathcal{S}((I + i\alpha_1 A_1) \otimes i(I + \alpha_2 A_2))]. \quad (\text{A8})$$

Combining this with equation (A5) completes the proof.

A.2. Proof of lemma 3

We first write down the tensor representation of the projective measurement, $I + \beta A_1 \otimes A_2$ for $\beta = \pm 1$.

$$\begin{aligned} \langle e_i e_j | \mathcal{S}(I + \beta A_1 \otimes A_2) | e_k e_l \rangle &= \text{Tr} (e_i \otimes e_j (I + \beta A_1 \otimes A_2) e_k \otimes e_l (I + \beta A_1 \otimes A_2)) \\ &= a_{1,ijkl} + \beta (a_{6,ijkl} + a_{7,ijkl}) + a_{16,ijkl} \end{aligned} \quad (\text{A9})$$

Similarly to the previous proof, observe that,

$$\left\{ \begin{array}{l} \{+1, +1\} - \{+1, -1\} \\ = 2(a_4 + a_5) + 2(a_6 + a_7) + 2(a_8 + a_9) + 2(a_{12} + a_{13}), \\ \{-1, +1\} - \{-1, -1\} \\ = 2(a_4 + a_5) - 2(a_6 + a_7) - 2(a_8 + a_9) + 2(a_{12} + a_{13}), \\ \{+i, +i\} - \{+i, -i\} \\ = 2i(a_4 - a_5) - 2(a_6 + a_7) + 2(a_8 + a_9) + 2i(a_{12} - a_{13}), \\ \{-i, +i\} - \{-i, -i\} \\ = 2i(a_4 - a_5) + 2(a_6 + a_7) - 2(a_8 + a_9) + 2i(a_{12} - a_{13}), \end{array} \right. \quad (\text{A10})$$

We can solve the above for $i(a_6 + a_7)$, and obtain,

$$8(a_6 + a_7) = \{+1, +1\} - \{+1, -1\} - \{-1, +1\} + \{-1, -1\} \\ - \{+i, +i\} + \{+i, -i\} + \{-i, +i\} - \{-i, -i\} \quad (\text{A11})$$

$$= \sum_{\alpha \in \{\pm 1\}^2} \alpha_1 \alpha_2 [\mathcal{S}((I + \alpha_1 A_1) \otimes (I + \alpha_2 A_2)) - \mathcal{S}((I + i\alpha_1 A_1) \otimes (I + i\alpha_2 A_2))]. \quad (\text{A12})$$

A.3. Relation with reference [20]

Bravyi *et al* has considered to remove k qubits in a given $n + k$ -qubit circuit at the cost of $O(kd^k)$ classical computation, where d defined to be proportional to the number of gates applied to the k -qubit system. The technique utilized in their work, in particular, figure 2 in reference [20] can also provide a derivation to the above lemmas when combined with our recent technique developed in reference [28].

Appendix B. Proof of lemma 2

Suppose that we are applying $\mathcal{S}(e^{i\theta A_1 \otimes A_2})$ to some state ρ and want to decompose the gate. We name each operation in the decomposition as,

$$\begin{aligned} \Phi_{1,\beta} &= \mathcal{S}(I \otimes I), \\ \Phi_{2,\beta} &= \mathcal{S}(A_1 \otimes A_2), \\ \Phi_{3,\beta} &= \beta \mathcal{M}_{A_1,\beta} \otimes \mathcal{S}(e^{i\pi A_2/4}), \\ \Phi_{4,\beta} &= \beta \mathcal{M}_{A_1,\beta} \otimes \mathcal{S}(e^{-i\pi A_2/4}), \\ \Phi_{5,\beta} &= \beta \mathcal{S}(e^{i\pi A_1/4}) \otimes \mathcal{M}_{A_2,\beta}, \\ \Phi_{6,\beta} &= \beta \mathcal{S}(e^{-i\pi A_1/4}) \otimes \mathcal{M}_{A_2,\beta}. \end{aligned} \quad (\text{B1})$$

which is not physical when $\beta_{3,4,5,6} = -1$ but achievable with classical post processing. $\mathcal{M}_{A_i,\beta}$ is a postselective measurement operation, which has been introduced in the main text. For convenience, we define coefficients $\{c_i\}_{i=1}^6$ as

$$\begin{aligned} c_1 &= \cos^2 \theta, \\ c_2 &= \sin^2 \theta, \\ c_3 &= -c_4 = c_5 = -c_6 = \cos \theta \sin \theta, \end{aligned} \quad (\text{B2})$$

Then,

$$\mathcal{S}(e^{i\theta A_1 \otimes A_2})|\rho\rangle\rangle \quad (\text{B3})$$

$$= \left[c_1 \Phi_{1,\beta} + c_2 \Phi_{2,\beta} \right. \quad (\text{B4})$$

$$\left. \sum_{\beta \in \{1, -1\}} \text{Tr} \left(\rho \frac{I + \beta A_1}{2} \right) (c_3 \Phi_{3,\beta} + c_4 \Phi_{4,\beta}) \right] \quad (\text{B5})$$

$$\sum_{\beta \in \{1, -1\}} \text{Tr} \left(\rho \frac{I + \beta A_2}{2} \right) (c_5 \Phi_{5,\beta} + c_6 \Phi_{6,\beta}) \Big] |\rho\rangle\rangle \quad (\text{B6})$$

We take a naive algorithm to bound the error of the decomposition. We define a probabilistic map below

$$\begin{aligned} \Phi_1 &= \mathcal{S}(I \otimes I), \\ \Phi_2 &= \mathcal{S}(A_1 \otimes A_2), \\ \Phi_3 &= \mathcal{M}'_{A_1} \otimes \mathcal{S}(e^{i\pi A_2/4}), \\ \Phi_4 &= \mathcal{M}'_{A_1} \otimes \mathcal{S}(e^{-i\pi A_2/4}), \\ \Phi_5 &= \mathcal{S}(e^{i\pi A_1/4}) \otimes \mathcal{M}'_{A_2}, \\ \Phi_6 &= \mathcal{S}(e^{-i\pi A_1/4}) \otimes \mathcal{M}'_{A_2}, \end{aligned} \quad (\text{B7})$$

where \mathcal{M}'_{A_i} acts on a state ρ probabilistically as,

$$\mathcal{M}'_{A_i}(\rho) \rightarrow b \mathcal{M}_{A_i,b}(\rho) \quad (\text{B8})$$

where b is a random variable with probability distribution $p(b = \pm 1) = \text{Tr}(\rho \frac{I \pm A_i}{2})$. Again, when $b = -1$ this map is non-physical but can be realized with classical post processing. Φ_i becomes $\Phi_{i,b}$ with probability $\text{Tr}(\rho \frac{I \pm A_i}{2})$, and therefore,

$$\mathbb{E}[\Phi_i] = \text{Tr} \left(\rho \frac{I + A_1}{2} \right) \Phi_{i,+1} + \text{Tr} \left(\rho \frac{I - A_1}{2} \right) \Phi_{i,-1} \quad (\text{B9})$$

for $i = 3, 4$. A similar equality holds for $i = 5, 6$. This yields,

$$\mathcal{S}(e^{i\theta A_1 \otimes A_2})|\rho\rangle\rangle = \sum_{i=1}^6 c_i \mathbb{E}[\Phi_i|\rho\rangle\rangle] \quad (\text{B10})$$

Suppose that we take N samples for each $i = 1, \dots, 6$ to estimate $\mathbb{E}[\Phi_i|\rho\rangle\rangle]$. The $i = 1, 2$ cases are not probabilistic and hence do not introduce error. We are left to consider the error induced by $i = 3, 4, 5, 6$. In this case, we can estimate $|\mu_i\rangle\rangle = \mathbb{E}[\Phi_i|\rho\rangle\rangle]$ by

$$|\bar{\mu}_i\rangle\rangle = \frac{1}{N} \sum_{j=1}^N \Phi_{i,b_{ij}}|\rho\rangle\rangle, \quad (\text{B11})$$

where $\{b_{ij}\}_{j=1}^N$ are samples drawn from the distribution which is identical to the above mentioned b . Now the difference between the true state $\mathcal{S}(e^{i\theta A_1 \otimes A_2})|\rho\rangle\rangle$ and the estimated $\sum_{i=1}^6 c_i |\bar{\mu}_i\rangle\rangle$ is,

$$\begin{aligned} \mathcal{S}(e^{i\theta A_1 \otimes A_2})|\rho\rangle\rangle - \sum_{i=1}^6 c_i |\bar{\mu}_i\rangle\rangle &= \left(\text{Tr} \left(\rho \frac{I + A_1}{2} \right) - \frac{1}{2N} \sum_{j=1}^N (b_{3j} + 1) \right) c_3 \Phi_{3,+1}|\rho\rangle\rangle \\ &\quad + \left(\text{Tr} \left(\rho \frac{I - A_1}{2} \right) - \frac{1}{2N} \sum_{j=1}^N (1 - b_{3j}) \right) c_3 \Phi_{3,-1}|\rho\rangle\rangle \\ &\quad + \left(\text{Tr} \left(\rho \frac{I + A_1}{2} \right) - \frac{1}{2N} \sum_{j=1}^N (b_{4j} + 1) \right) c_4 \Phi_{4,+1}|\rho\rangle\rangle \\ &\quad + \left(\text{Tr} \left(\rho \frac{I - A_1}{2} \right) - \frac{1}{2N} \sum_{j=1}^N (1 - b_{4j}) \right) c_4 \Phi_{4,-1}|\rho\rangle\rangle \\ &\quad + \left(\text{Tr} \left(\rho \frac{I + A_2}{2} \right) - \frac{1}{2N} \sum_{j=1}^N (b_{5j} + 1) \right) c_5 \Phi_{5,+1}|\rho\rangle\rangle \\ &\quad + \left(\text{Tr} \left(\rho \frac{I - A_2}{2} \right) - \frac{1}{2N} \sum_{j=1}^N (1 - b_{5j}) \right) c_5 \Phi_{5,-1}|\rho\rangle\rangle \end{aligned}$$

$$\begin{aligned}
& + \left(\text{Tr} \left(\rho \frac{I + A_2}{2} \right) - \frac{1}{2N} \sum_{j=1}^N (b_{6j} + 1) \right) c_6 \Phi_{6,+1} |\rho\rangle\rangle \\
& + \left(\text{Tr} \left(\rho \frac{I - A_2}{2} \right) - \frac{1}{2N} \sum_{j=1}^N (1 - b_{5j}) \right) c_6 \Phi_{6,-1} |\rho\rangle\rangle. \quad (\text{B12})
\end{aligned}$$

$\frac{1 \pm b_{ij}}{2}$ is a Bernoulli random variable with the expectation $\frac{I \pm A_1}{2}$ and $\frac{I \pm A_2}{2}$ respectively for $i = 3, 4$ and $i = 5, 6$. This means that, for example, the difference between $\frac{1}{2N} \sum_{j=1}^N (b_{3j} + 1)$ and $\frac{I + A_1}{2}$ is bounded by $\epsilon > 0$, that is, $\left| \frac{1}{2N} \sum_{j=1}^N (b_{3j} + 1) - \text{Tr} \left[\rho \frac{I + A_1}{2} \right] \right| \leq \epsilon$ with probability at most $1 - \exp(-2\epsilon^2 N)$ from Hoeffding's inequality. The same bound holds for every term in equation (B12). Noting that if $\left| \frac{1}{2N} \sum_{j=1}^N (b_{3j} + 1) - \text{Tr} \left[\rho \frac{I + A_1}{2} \right] \right| \leq \epsilon$ holds, $\left| \frac{1}{2N} \sum_{j=1}^N (1 - b_{3j}) - \text{Tr} \left[\rho \frac{I - A_1}{2} \right] \right| \leq \epsilon$ also holds, the probability that at least one of the differences in equation (B12) is larger than ϵ is at most $4 \exp(-2\epsilon^2 N)$, by union bound. Therefore, with probability at least $1 - 4 \exp(-2\epsilon^2 N)$,

$$\begin{aligned}
\left\| \mathcal{S}(e^{i\theta A_1 \otimes A_2}) |\rho\rangle\rangle - \sum_{i=1}^6 c_i |\bar{\mu}_i\rangle\rangle \right\| & \leq \epsilon \left\| \sum_{i=3}^6 \sum_{\beta \in \{1, -1\}} c_i \Phi_{i,\beta} |\rho\rangle\rangle \right\| \\
& \leq \epsilon \sum_{i=3}^6 \sum_{\beta \in \{1, -1\}} \|c_i \Phi_{i,\beta} |\rho\rangle\rangle\|, \quad (\text{B13})
\end{aligned}$$

holds for any norm $\|\cdot\|$. The second inequality follows from the triangle inequality. Considering the trace norm $\|\cdot\|_1$, which gives $\|\Phi_{i,\beta} |\rho\rangle\rangle\|_1 = 1$, and taking $|c_i| \leq 1$ into account, we get

$$\left\| \mathcal{S}(e^{i\theta A_1 \otimes A_2}) |\rho\rangle\rangle - |\bar{\mu}_i\rangle\rangle \right\| \leq 8\epsilon. \quad (\text{B14})$$

With this, we conclude that, given the desired error $1/\epsilon$ and a probability $1 - \delta$ by which we wish to lower-bound the probability of getting the error larger than ϵ , we can take $N = -\frac{32}{\epsilon^2} \ln(1 - \delta)$.

Appendix C. Time-like cut for identity channel

The time-like cut approach proposed in reference [16] can be derived in the following manner. Let us consider an identity channel \mathcal{I}_a on the a th qubit. It can be expanded as,

$$\mathcal{I}_a = \sum_{i_a=0}^3 \sum_{j_a=0}^3 |e_{i_a}\rangle\rangle \langle\langle e_{j_a}| \langle\langle e_{i_a} | \mathcal{I}_a | e_{j_a}\rangle\rangle. \quad (\text{C1})$$

Since we assumed $|e_i\rangle\rangle$ are orthonormal to each other and $\mathcal{I}|\rho\rangle\rangle = |\rho\rangle\rangle$ for any ρ ,

$$\mathcal{I}_a = \sum_{i_a=0}^3 |e_{i_a}\rangle\rangle \langle\langle e_{i_a}| \quad (\text{C2})$$

If we apply this to a n -qubit density matrix $|\rho\rangle\rangle = \sum_{j_1, \dots, j_n} \rho_j |e_{j_1} e_{j_2} \dots e_{j_n}\rangle\rangle$ in this form, we see,

$$\mathcal{I}|\rho\rangle\rangle = \sum_{i_a=0}^3 |e_{i_a}\rangle\rangle \langle\langle e_{i_a} | \rho\rangle\rangle \quad (\text{C3})$$

$$= \sum_{i_a=0}^3 \text{Tr}_a(e_{i_a} \rho) \otimes |e_{i_a}\rangle\rangle. \quad (\text{C4})$$

Choosing $|e_{j_1}\rangle\rangle$ to be Pauli matrices $\{I, X, Y, Z\}/\sqrt{2}$, we conclude,

$$\mathcal{I}|\rho\rangle\rangle = \frac{1}{2} \sum_{A \in \{I, X, Y, Z\}} \text{Tr}_a(A_a \rho) \otimes A_a. \quad (\text{C5})$$

This equation implies that we can first measure expectation values of X, Y, Z at the a th qubit and then re-input each eigenstates.

Appendix D. Analysis of the simple example given in section 3.1

D.1. Cost of time-like cut

First, we consider the ‘time-like’ cut approach. Let us name two-qubit on which the CZ gate acts a and b , and let σ_a and σ_b be their 1-qubit reduced density matrices after the gate $U_1 \otimes V_1$.

We assume,

- One n -qubit device is available.
- Qubits that are measured in the basis of an observable O_i can be reused to prepare the input state ρ_j .

We take the following naive approach to estimate $\langle O_f \rangle$. First, divide the allowed number of circuit runs N into $N/2$ to run the divided circuit for equal times. $N/2$ runs are further divided into $N/128$ runs to run the circuit with O_i and ρ_j for $i, j \in \{1, 2, \dots, 8\}^2$ [30]. With each $N/128$ runs with a pair (O_i, ρ_j) , we estimate the value of the tensor network below. Since we assumed O_f is a tensor product of Pauli Z 's and O_i is drawn from I, X, Y, Z , from each run we obtain a measurement result $o_{ij,r} = \pm 1$, where r is the index to distinguish the runs. Using $o_{ij,r}$, we estimate the above tensor network by $\tilde{o}_{ij} = \frac{1}{N/128} \sum_{r=1}^{N/128} o_{ij,r}$. Since $o_{ij,r}$ is a random variable which takes $\{+1, -1\}$, this estimator \tilde{o}_{ij} approximately follows a normal distribution with an expectation $\mathbb{E}[o_{ij}]$ and a variance $\frac{1}{N/128}(1 - \mathbb{E}[o_{ij}]^2) \leq \frac{1}{N/128}$, for sufficiently large N . Therefore, with $N/128$ runs of a quantum circuit, we can estimate the value of the above tensor network with the variance $\frac{128}{N}$ at most. We can obtain the same result for the other cluster.

For each i, j, k, l , the tensor in the sum of figure 4 is estimated by the product of the above estimators because the measurement result is independent on each cluster. The variance of each tensor network can be evaluated because they are a product of two random variables approximately drawn from normal distribution with variance at most $\frac{128}{N}$, and it is at most $\frac{64}{N}$. We have that $c_{ijkl} \in \{\pm 1/2^4\}$. This reduces the variance of each term in the summation to $1/(2N)$. However, when we take the sum since each term can be approximated by a normal distribution with the variance at most $1/(2N)$ and we take summation of $8^4 = 4096$ terms, the result has the variance at most $2048/N$.

D.2. Cost of space-like cut

For the space-like cut, we divide the allowed number of circuit runs, N , to $N/6$ [31]. First, we run eight circuits that do not involve the measurement in the middle and obtain estimators for these four tensor network. Each of the estimators has the variance at most $6/N$. Let us now move on to the circuits with the measurement. For arbitrary density matrix ρ' , the Z measurement produces the density matrix $\left(\frac{1+\alpha}{2}\right) \rho' \left(\frac{1+\alpha}{2}\right) / p_\alpha$ with probability p_α . Therefore, to obtain the above decomposition, we need to know the normalization factor p_α . With $N/6$ circuit runs, p_α is estimated to the variance $p_\alpha(1 - p_\alpha)/(N/6)$ which is at most $6/(4N) = 3/(2N)$. Note that when N is large, the distribution of the estimator \tilde{p}_α can be thought of as a normal distribution. Conditioned on α , we construct an estimator of O_f . Since α is obtained with probability p_α , for each α we have $p_\alpha N/6$ samples to estimate O_f . Therefore, for each α , the estimator of O_f has the variance of $6/p_\alpha N$. Since the estimator of the tensor network is the product of the estimators of conditioned O_f and p_α , its variance is $\frac{6}{N} \frac{1}{p_\alpha + 1/[p_\alpha(1-p_\alpha)]} \leq \frac{3}{2N}$. Each pair of the tensor network in figure 2 is multiplied together, and if we perform this with the estimators obtained above, the variance of each term is at most $\frac{3}{N}$. We further multiply each term with $\pm 1/2$, then the variance is reduced to $\frac{3}{4N}$. Finally, the summation of 10 such term leads to the variance of $\frac{15}{2N}$.

Appendix E. Proof of theorem 5

We follow the approach taken in reference [16]. The task here is to perform the decomposition of m -qubit circuit like the one shown in figure 5 so that the original quantum circuit can be approximated with an n -qubit quantum computer, where the input state ρ is initialized in $|0\rangle\langle 0|^{\otimes m}$ and O_f is an output (diagonal) observable calculated from some output function $f: \{0, 1\}^m \rightarrow [-1, 1]$. We want to estimate $\mathbb{E}_y[f(y)]$ for n -bit measurement outcomes y to some accuracy $\epsilon > 0$ with some high probability $1 - \delta$.

Let the number of space-like cuts and time-like cuts performed in the decomposition be M_s and M_t respectively. We assume the space-like cuts are performed only on CZ gates. We redefine the probabilistic map Φ'_i that is used to decompose $\mathcal{S}(e^{i\theta A \otimes B})$ as,

$$\Phi'_1 = \mathcal{S}(I \otimes I),$$

$$\Phi'_2 = \mathcal{S}(A \otimes B),$$

$$\Phi'_3 = \mathcal{M}_A \otimes \mathcal{S}(e^{i\pi B/4}),$$

$$\begin{aligned}
\Phi'_4 &= \mathcal{M}_A \otimes \mathcal{S}(e^{-i\pi B/4}), \\
\Phi'_5 &= \mathcal{S}(e^{i\pi A/4}) \otimes \mathcal{M}_B, \\
\Phi'_6 &= \mathcal{S}(e^{-i\pi A/4}) \otimes \mathcal{M}_B,
\end{aligned} \tag{E1}$$

where \mathcal{M}_A and \mathcal{M}_B and the projective measurement of A and B . Let $s_k \in \{1, \dots, 6\}$ be an index of the above probabilistic map Φ'_{s_k} applied to the k th space-like cut $k \in \{1, \dots, M_s\}$ and t_l be an index of an observable-state pair (O_{t_l}, ρ_{t_l}) in equation (11) applied to the l th time-like cut $l \in \{1, \dots, M_t\}$. The coefficients associated with a space-like cut (equation (B2)) and a time-like cut (equation (11)) are redefined as $c_{s_k}^{\text{space}}$ and $c_{t_l}^{\text{time}}$, respectively. With one set of indices, $s = \{s_k\}_{k=1}^{M_s} \in \{1, \dots, 6\}^{M_s}$ and $t = \{t_l\}_{l=1}^{M_t} \in \{1, \dots, 8\}^{M_t}$, we can define a corresponding quantum circuit which is induced by replacing every cut two-qubit gate by Φ'_{s_k} and every cut qubit line by the measurement of O_{t_l} and the preparation of ρ_{t_l} .

When we run this circuit on n -qubit quantum device, we get the measurement outcomes at each cut, which is a string of ± 1 from Φ'_{s_k} and the measurement of O_{t_l} , and the ones at the output qubit which is a bitstring of length n . Let such outcomes from the k th space-like cut, the l th time-like cut and the output qubits be $b_{s_k}^{\text{space}} \in \{+1, -1\}$, $b_{t_l}^{\text{time}} \in \{+1, -1\}$ and $y_{(s,t)} \in \{0, 1\}^n$, respectively. Since $s_k = 1, 2$ does not involve measurement, we define $b_1^{\text{space}} = b_2^{\text{space}} = 1$. With the definition above and the equality for performing the decomposition (equations (B10) and (11), figures 1 and 3), notice that,

$$\mathbb{E}_y[f(y)] = \sum_{s \in \{1, \dots, 6\}^{M_s}} \sum_{t \in \{1, \dots, 8\}^{M_t}} \prod_{k=1}^{M_s} c_{s_k} \prod_{l=1}^{M_t} c_{t_l} \mathbb{E}_{\{b_{s_k}^{\text{space}}\}, \{b_{t_l}^{\text{time}}\}, y_{(s,t)}} \left[\prod_{k=1}^{M_s} b_{s_k}^{\text{space}} \prod_{l=1}^{M_t} b_{t_l}^{\text{time}} f(y_{(s,t)}) \right], \tag{E2}$$

where the expectation on the right-hand side is defined over a distribution of $\{b_{s_k}^{\text{space}}\}$, $\{b_{t_l}^{\text{time}}\}$ and $y_{(s,t)}$ for a quantum circuit induced by a given set of indices (s, t) .

We can take a Monte-Carlo approach to estimate the sum of the right-hand side of equation (E2). If we sample s and t from a uniform distribution on $\{1, \dots, 6\}^{M_s}$ and $\{1, \dots, 8\}^{M_t}$ respectively, equation (E2) can be rewritten as,

$$\mathbb{E}_y[f(y)] = \mathbb{E}_{(s,t), \{b_{s_k}^{\text{space}}\}, \{b_{t_l}^{\text{time}}\}, y_{(s,t)}} \left[6^{M_s} 8^{M_t} \prod_{k=1}^{M_s} c_{s_k} b_{s_k}^{\text{space}} \prod_{l=1}^{M_t} c_{t_l} b_{t_l}^{\text{time}} f(y_{(s,t)}) \right], \tag{E3}$$

Let us define a random variable

$$X_{(s,t)} = 6^{M_s} 8^{M_t} \prod_{k=1}^{M_s} c_{s_k} b_{s_k}^{\text{space}} \prod_{l=1}^{M_t} c_{t_l} b_{t_l}^{\text{time}} f(y_{(s,t)}). \tag{E4}$$

Let $(s^{(i)}, t^{(i)})_{i=1}^N$ be N randomly sampled (s, t) pair. Then, $\mathbb{E}_y[f(y)]$ can be estimated by $\frac{1}{N} \sum_{i=1}^N X_{(s^{(i)}, t^{(i)})}$. We will use the Hoeffding's inequality to bound the error of this Monte-Carlo approach. The magnitude of $X_{(s,t)}$ is bounded by,

$$\left| 6^{M_s} 8^{M_t} \prod_{k=1}^{M_s} c_{s_k} b_{s_k}^{\text{space}} \prod_{l=1}^{M_t} c_{t_l} b_{t_l}^{\text{time}} f(y_{(s,t)}) \right| \leq 3^{M_s} 4^{M_t}, \tag{E5}$$

because $|c_{t_l}| = 1/2$, $|f(y_{(s,t)})| \leq 1$, $|b_{s_k}^{\text{space}}| = 1$, $|b_{t_l}^{\text{time}}| = 1$, and $|c_{s_k}| = 1/2$ which follows from the assumption that the space-like cuts are performed only on CZ gates. With the above bound of the magnitude, the Hoeffding's inequality guarantees that,

$$\Pr \left[\left| \frac{1}{N} \sum_{i=1}^N X_{(s^{(i)}, t^{(i)})} - \mathbb{E}_y[f(y)] \right| \geq \epsilon \right] \tag{E6}$$

$$\geq 1 - 2 \exp \left(- \frac{N \epsilon^2}{2 \cdot 9^{M_s} \cdot 16^{M_t}} \right). \tag{E7}$$

Therefore, for given ϵ and the probability $1 - \delta$ to which we want to bound the probability of getting an error larger than ϵ , we take $N = \frac{2 \cdot 9^{M_s} \cdot 16^{M_t}}{\epsilon^2} \ln \left(\frac{1}{2\delta} \right)$.

ORCID iDs

Kosuke Mitarai  <https://orcid.org/0000-0002-9056-316X>

References

- [1] Barends R *et al* 2014 *Nature* **508** 500
- [2] Bernien H *et al* 2017 *Nature* **551** 579
- [3] Wright K *et al* 2019 arXiv:1903.08181
- [4] Preskill J 2018 *Quantum* **2** 79
- [5] Harrow A W and Montanaro A 2017 *Nature* **549** 203
- [6] Boixo S, Isakov S V, Smelyanskiy V N, Babbush R, Ding N, Jiang Z, Bremner M J, Martinis J M and Neven H 2018 *Nat. Phys.* **14** 595
- [7] Neill C *et al* 2018 *Science* **360** 195
- [8] Bravyi S, Gosset D and König R 2018 *Science* **362** 308
- [9] Peruzzo A, McClean J, Shadbolt P, Yung M-H, Zhou X-Q, Love P J, Aspuru-Guzik A and O'Brien J L 2014 *Nat. Commun.* **5** 4213
- [10] Kandala A, Mezzacapo A, Temme K, Takita M, Brink M, Chow J M and Gambetta J M 2017 *Nature* **549** 242
- [11] Nam Y *et al* 2019 arXiv:1902.10171
- [12] Farhi E, Goldstone J and Gutmann S 2014 arXiv:1411.4028
- [13] Otterbach J S *et al* 2017 arXiv:1712.05771
- [14] Mitarai K, Negoro M, Kitagawa M and Fujii K 2018 *Phys. Rev. A* **98** 032309
- [15] Havlíček V, Córcoles A D, Temme K, Harrow A W, Kandala A, Chow J M and Gambetta J M 2019 *Nature* **567** 209
- [16] Peng T, Harrow A, Ozols M and Wu X 2019 arXiv:1904.00102
- [17] Mitarai K and Fujii K 2019 *Phys. Rev. Res.* **1** 013006
- [18] Chen Z-Y, Zhou Q, Xue C, Yang X, Guo G-C and Guo G-P 2018 *Sci. Bull.* **63** 964
- [19] Pednault E, Gunnels J A, Nannicini G, Horesh L, Magerlein T, Solomonik E, Draeger E W, Holland E T and Wisnieff R 2017 arXiv:1710.05867
- [20] Bravyi S, Smith G and Smolin J A 2016 *Phys. Rev. X* **6** 021043
- [21] Barenco A, Bennett C H, Cleve R, DiVincenzo D P, Margolus N, Shor P, Sleator T, Smolin J A and Weinfurter H 1995 *Phys. Rev. A* **52** 3457
- [22] Vidal G 2000 *J. Mod. Opt.* **47** 355–76
- [23] Shi Y-Y, Duan L-M and Vidal G 2006 *Phys. Rev. A* **74** 022320
- [24] Markov I L and Shi Y 2008 *SIAM J. Comput.* **38** 963
- [25] Vidal G 2003 *Phys. Rev. Lett.* **91** 147902
- [26] Endo S, Benjamin S C and Li Y 2018 *Phys. Rev. X* **8** 031027
- [27] Wecker D, Hastings M B and Troyer M 2015 *Phys. Rev. A* **92** 042303
- [28] Mitarai K, Yan T and Fujii K 2019 *Phys. Rev. Appl.* **11** 044087
- [29] Biamonte J and Bergholm V 2017 Tensor networks in a nutshell (arXiv:1708.00006 [quant-ph])
- [30] Note that this approach is almost equivalent for a large N to the approach proposed in reference [16], which proceeds by drawing the pair i, j from uniform distribution.
- [31] This is also a naive approach and there is a possibility to optimize the way of division, as mentioned in the main text.