

Computationally Efficient Quantum Expectation with Extended Bell Measurements

Ruho Kondo¹, Yuki Sato¹, Satoshi Koide¹, Seiji Kajita¹, and Hideki Takamatsu²

¹Toyota Central R&D Labs., Inc., 41-1, Yokomichi, Nagakute, Aichi 480-1192, Japan

²Toyota Motor Corporation, 1 Toyota-Cho, Toyota, Aichi 471-8571, Japan

Evaluating an expectation value of an arbitrary observable $A \in \mathbb{C}^{2^n \times 2^n}$ through naïve Pauli measurements requires a large number of terms to be evaluated. We approach this issue using a method based on Bell measurement, which we refer to as the extended Bell measurement method. This analytical method quickly assembles the 4^n matrix elements into at most 2^{n+1} groups for simultaneous measurements in $O(nd)$ time, where d is the number of non-zero elements of A . The number of groups is particularly small when A is a band matrix. When the bandwidth of A is $k = O(n^c)$, the number of groups for simultaneous measurement reduces to $O(n^{c+1})$. In addition, when non-zero elements densely fill the band, the variance is $O((n^{c+1}/2^n) \text{tr}(A^2))$, which is small compared with the variances of existing methods. The proposed method requires a few additional gates for each measurement, namely one Hadamard gate, one phase gate and at most $n - 1$ CNOT gates. Experimental results on an IBM-Q system show the computational efficiency and scalability of the proposed scheme, compared with existing state-of-the-art approaches. Code is available at <https://github.com/ToyotaCRDL/extended-bell-measurements>.

1 Introduction

Variational quantum algorithms (VQAs) [1, 2] have attracted attention in recent years because they are suitable for so-called Noisy Intermediate-Scale Quantum (NISQ) computers [3], which do not have a fault-tolerant mechanism and have only tens or hundreds of qubits.

Ruho Kondo: r-kondo@mosk.tytlabs.co.jp

The VQA was first introduced in the field of chemistry but it has wide application, such as in solving linear systems [4–8], optimization [9–11], and machine learning [12–18]. In VQAs, the quantum state, $|\psi\rangle$, is parametrized by learning parameters held on a classical computer, and is used to calculate the expectation value, $\langle\psi|A|\psi\rangle$, whose A depends on the task. The desired quantum state is then obtained by minimizing the expectation value using a classical optimization algorithm.

The bottleneck of a VQA relates to the computational cost of calculating the expectation value [19]. A popular method of evaluating the expectation value is a Pauli measurement that writes A in the Pauli basis; i.e., $A = \sum_{i=1}^{\mathcal{N}} c_i \mathbf{P}_i$, where $\mathbf{P}_i$ is the tensor product of Pauli matrices and $c_i = \text{Tr}(A^\dagger \mathbf{P}_i)/2^n$ is the corresponding coefficient, with $(\bullet)^\dagger$ indicating the Hermitian conjugate. The drawback of a naïve Pauli measurement is that it requires a large number of unique circuits to evaluate $\langle\psi|A|\psi\rangle = \sum_{i=1}^{\mathcal{N}} c_i \langle\psi|\mathbf{P}_i|\psi\rangle$ where $\mathcal{N}$ reaches 4^n when A is a fully dense matrix.

Reducing the number of unique circuits for Pauli measurements is an interesting topic relating to the VQA [20–26]. Both qubit-wise commuting (QWC) [21, 27, 28] and general commuting (GC) [24] methods, and their variants [22, 23, 26], reduce the number of unique circuits on the basis that commutative operators can be simultaneously diagonalized with the same unitary; i.e., commutative Pauli bases are simultaneously measurable. This approach successfully reduces the number of unique circuits but its weakness is that finding the smallest groups for simultaneous measurements itself is an NP-hard problem. These methods approximate solutions that are unfortunately not optimal in general. In other words, a smaller number of unique circuits results in a longer calculation time, and a shorter calcu-

lation time results in a larger number of unique circuits.

Another problem is that the expression of general matrices in the Pauli basis is numerically inefficient. Suppose a sparse matrix A that has a few non-zero elements. In the extreme case, for example, if there exists just one non-zero element in A , then 2^n Pauli strings are required to represent A . Despite the inefficiency of the Pauli basis, the sparse matrix is widely used in the domain of computer science. A typical application is the finite element method (FEM) [29] used for structural analysis, fluid dynamics, heat transfer, electromagnetic analysis and many other continuum mechanics. In the FEM, three dimensional real-space is discretized by a finite number of elements and the stiffness of the system is constructed by the nodal coordinates and properties of each element. The resulting global stiffness matrix, A , in an appropriate node number ordering represents a band matrix because only the neighboring nodes in the elements can interact with each other. Then, solving $Ax = b$, where b denotes the external inputs, gives the desired response x . Solving $Ax = b$ using VQAs has long been studied [4–8] but most of the VQAs limit the form of A . For example, the assumption is that A can be represented as the sum of a few weighted unitaries or weighted Pauli strings [4–6] or can be decomposed into a few measurable terms using the problem-specific conditions [7, 8]. Recently, an efficient expectation estimation method for a sparse Hamiltonian has been developed [30], but it remains unsuitable for the NISQ device because of its two-qubit entangling gate counts. Efficient expectation evaluations for the non-Pauli observables in terms of both the number of measurements and the circuit complexity are in great demand, such as for FEM applications, but the above issues have not yet been resolved.

A completely different approach that can be taken to reduce the computational cost for the expectation evaluation is to use the so-called classical shadow [31–33], which reduces the number of measurements for the Pauli observables. However, in the case of non-Pauli observables, this approach is not suitable for NISQ devices because it requires $O(n^2/\log(n))$ two-qubit entangling gates as pointed out in the original literature [31].

In this paper, we propose an alternative

method that we refer to as the *extended Bell measurement (XBM)*. This method is based on not Pauli measurements but Bell measurements. The basic idea is inspired by Ref. [7], where the expectation value required to solve the Poisson equation was evaluated using Bell measurements. We generalize this method for an arbitrary expectation evaluation. In addition, we derive that the present method is closely related to the classical shadow with random Clifford measurements [31]. The XBM method has the following advantages over existing methods.

1. Fast grouping: The XBM method does not solve any difficult NP-hard problems, but it analytically groups the simultaneous measurements.
2. Small number of circuits: The method requires the number of unique circuits to be of the same order as the empirical number of an existing state-of-the-art Pauli measurement-based method, such as the GC method.
3. Fewer two-qubit entangling gates: The method takes at most $n - 1$ CNOT gates for each measurement operator.
4. Tight upper bound of the variance when A is the densely filled band matrix.

Concretely, when A is a band matrix whose bandwidth is $k = O(n^c)$, the number of distinct measurement operators is $O(n^{c+1})$. In addition, when the non-zero elements are densely filled in the band, the upper bound of the variance of the expectation evaluation is $O((n^{c+1}/2^n) \text{tr}(A^2))$, which is smaller than that of the classical shadow [31].

Our algorithm is implemented with Qiskit [34] and its code is available at <https://github.com/ToyotaCRDL/extended-bell-measurements>. Additionally, real-device experiments are carried out using an IBM-Q system. The results show that the accuracy of the expectation evaluation is comparable to that of naïve Pauli measurements and QWC but the execution time is shorter than the execution times of the existing methods when A is a band matrix.

The remainder of the paper is organized as follows: Section 2 describes the XBM method proposed in this paper. Section 3 presents algorithms of the XBM method and derives the

computational complexity. Section 4 presents the experimental results. Section 5 concludes the paper.

2 Method

2.1 Two-qubit Example of Proposed Method

Before describing the general proposed method, we show an example with two-qubit. Let $|\varphi\rangle \in \mathbb{C}^4$ be an arbitrary two-qubit quantum state. Consider the following expectation value:

$$\begin{aligned} \langle\varphi|A|\varphi\rangle &= \langle\varphi| \begin{bmatrix} A_{00} & A_{01} & A_{02} & A_{03} \\ A_{10} & A_{11} & A_{12} & A_{13} \\ A_{20} & A_{21} & A_{22} & A_{23} \\ A_{30} & A_{31} & A_{32} & A_{33} \end{bmatrix} |\varphi\rangle \\ &= \sum_{b=0}^3 \sum_{c=0}^3 A_{bc} \langle\varphi|b\rangle \langle c|\varphi\rangle. \end{aligned}$$

The diagonal components of A can be easily evaluated as

$$\sum_{b=0}^3 A_{bb} \langle\varphi|b\rangle \langle b|\varphi\rangle = \sum_{b=0}^3 A_{bb} |\langle b|\varphi\rangle|^2$$

where $|\langle b|\varphi\rangle|^2$ is the probability of obtaining a bit-string b when measuring the quantum state $|\varphi\rangle$ in the computational basis. Meanwhile, evaluating the off-diagonal parts of A is not trivial. Consider A_{12} for example. The expectation value of A_{12} can be rearranged as

$$\begin{aligned} \langle\varphi| \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & A_{12} & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} |\varphi\rangle &= \langle\varphi| (A_{12}|1\rangle\langle 2|) |\varphi\rangle \\ &= A_{12} (\text{Re}(\langle\varphi|1\rangle\langle 2|\varphi\rangle) + i\text{Im}(\langle\varphi|1\rangle\langle 2|\varphi\rangle)) \end{aligned}$$

where $\text{Re}(\bullet)$ and $\text{Im}(\bullet)$ are the real and imaginary parts of $(\bullet)$, respectively. The real part of $\langle\varphi|1\rangle\langle 2|\varphi\rangle$ is further rearranged as

$$\begin{aligned} &\text{Re}(\langle\varphi|1\rangle\langle 2|\varphi\rangle) \\ &= \frac{1}{2} \left(\left| \frac{\langle 01| + \langle 10|}{\sqrt{2}} |\varphi\rangle \right|^2 - \left| \frac{\langle 01| - \langle 10|}{\sqrt{2}} |\varphi\rangle \right|^2 \right) \end{aligned}$$

where $|01\rangle = |1\rangle$ and $|10\rangle = |2\rangle$, respectively. The quantum states $(|01\rangle \pm |10\rangle)/\sqrt{2}$ can be prepared using the Hadamard gate and CNOT gate as

$$\frac{|01\rangle + |10\rangle}{\sqrt{2}} = \text{CNOT}(1,0)\text{H}(1)|01\rangle$$

$$\frac{|01\rangle - |10\rangle}{\sqrt{2}} = \text{CNOT}(1,0)\text{H}(1)|11\rangle$$

where $\text{H}(j)$ denotes the Hadamard gate acting on j th qubit, and $\text{CNOT}(j,k)$ is the CNOT gate whose control and target qubits are j th and k th qubits, respectively. Here, the rightmost bit is 0th bit whereas the leftmost bit is $(n-1)$ th bit. Using these preparation unitaries, $\text{Re}(\langle\varphi|1\rangle\langle 2|\varphi\rangle)$ is written as

$$\begin{aligned} \text{Re}(\langle\varphi|1\rangle\langle 2|\varphi\rangle) &= \frac{1}{2} \left(\left| \langle 01|\text{H}(1)\text{CNOT}(1,0)|\varphi\rangle \right|^2 \right. \\ &\quad \left. - \left| \langle 11|\text{H}(1)\text{CNOT}(1,0)|\varphi\rangle \right|^2 \right). \end{aligned}$$

This value can be evaluated by the probabilities of obtaining bit-strings 01 and 11 when measuring $\text{H}(1)\text{CNOT}(1,0)|\varphi\rangle$ in the computational basis. In the same way, $\text{Im}(\langle\varphi|1\rangle\langle 2|\varphi\rangle)$ is written as

$$\begin{aligned} \text{Im}(\langle\varphi|1\rangle\langle 2|\varphi\rangle) &= \frac{1}{2} \left(\left| \langle 01|\text{H}(1)\text{CNOT}(1,0)\text{S}^\dagger(1)|\varphi\rangle \right|^2 \right. \\ &\quad \left. - \left| \langle 11|\text{H}(1)\text{CNOT}(1,0)\text{S}^\dagger(1)|\varphi\rangle \right|^2 \right) \end{aligned}$$

where $\text{S}(j)$ denotes the phase gate acting on j th qubit. This value can be evaluated by the probabilities of obtaining bit-strings 01 and 11 when measuring $\text{H}(1)\text{CNOT}(1,0)\text{S}^\dagger(1)|\varphi\rangle$ in the computational basis. In summary,

$$\begin{aligned} &\langle\varphi| \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & A_{12} & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} |\varphi\rangle \\ &= A_{12} \left[\frac{1}{2} \left(\left| \langle 01|M_{\text{Re}}^{(1,2)}|\varphi\rangle \right|^2 - \left| \langle 11|M_{\text{Re}}^{(1,2)}|\varphi\rangle \right|^2 \right) \right. \\ &\quad \left. + \frac{i}{2} \left(\left| \langle 01|M_{\text{Im}}^{(1,2)}|\varphi\rangle \right|^2 - \left| \langle 11|M_{\text{Im}}^{(1,2)}|\varphi\rangle \right|^2 \right) \right] \end{aligned}$$

where

$$M_{\text{Re}}^{(1,2)} := \text{H}(1)\text{CNOT}(1,0)$$

and

$$M_{\text{Im}}^{(1,2)} := H(1)\text{CNOT}(1,0)S^\dagger(1),$$

respectively. The remaining parts of A can be evaluated in the same manner.

The most important part of the proposed method is the simultaneous measurements. That is, some components of A can be evaluated using the same unitary. For example, A_{03} can be evaluated using the same unitaries as those for A_{12} , $H(1)\text{CNOT}(1,0)$ and $H(1)\text{CNOT}(1,0)S^\dagger(1)$, because

$$\begin{aligned} (|0\rangle + |3\rangle)/\sqrt{2} &= \text{CNOT}(1,0)H(1)|00\rangle, \\ (|0\rangle - |3\rangle)/\sqrt{2} &= \text{CNOT}(1,0)H(1)|10\rangle, \\ (|0\rangle + i|3\rangle)/\sqrt{2} &= S(1)\text{CNOT}(1,0)H(1)|00\rangle, \\ (|0\rangle - i|3\rangle)/\sqrt{2} &= S(1)\text{CNOT}(1,0)H(1)|10\rangle. \end{aligned}$$

In addition, A_{21} and A_{30} can be evaluated using the same unitaries as those for A_{12} and A_{03} , respectively, because $\text{Re}(\langle\varphi|b\rangle\langle c|\varphi\rangle) = \text{Re}(\langle\varphi|c\rangle\langle b|\varphi\rangle)$ and $\text{Im}(\langle\varphi|b\rangle\langle c|\varphi\rangle) = -\text{Im}(\langle\varphi|c\rangle\langle b|\varphi\rangle)$. Similarly, the real and imaginary parts of A_{01} , A_{10} , A_{23} and A_{32} can be evaluated using $H(0)$ and $H(0)S^\dagger(0)$, respectively, whereas the real and imaginary of A_{02} , A_{13} , A_{20} and A_{31} can be evaluated using $H(1)$ and $H(1)S^\dagger(1)$, respectively.

2.2 Expectation Estimation

We now generalize the construction of $\langle\varphi|A|\varphi\rangle$ for arbitrary number of qubits. Let $A \in \mathbb{C}^{2^n \times 2^n}$ be an arbitrary matrix and $|\varphi\rangle \in \mathbb{C}^{2^n}$ be an arbitrary n -qubit quantum state. The goal is to express $\langle\varphi|A|\varphi\rangle$ in the form $\mathbb{E}[\hat{a}] = \sum_i a_i P(\hat{a} = a_i)$, where $\hat{a}$ is a complex random variable, P is a probability distribution on the bit-strings obtained from a quantum circuit measurement, and a_i is an outcome, which is calculated from a measured bit-string and the components of A . The more general form, $\langle\psi_0|A|\psi_1\rangle$ (where $|\psi_0\rangle \neq |\psi_1\rangle$), can also be expressed in the same form (see Appendix A).

Here, we show a brief derivation of the closed-form expression of $\langle\varphi|A|\varphi\rangle$. In the subsequent paragraphs, we give an intuitive description of the derivation (see Appendix A for a detailed derivation). In this paper, the matrix A is expressed as

$$A = \sum_{b,c} A_{bc} |b\rangle\langle c| \quad (1)$$

where A_{bc} is the (b,c) -component of A . Using this representation, we give the closed-form expression of $\langle\varphi|A|\varphi\rangle$ derived in this study, where the notations appearing in Eq. (2) are defined in Table 1:

$$\begin{aligned} \langle\varphi|A|\varphi\rangle &= \sum_{b=0}^{2^n-1} A_{bb} |\langle b|\varphi\rangle|^2 + \sum_{b=0}^{2^n-1} \sum_{c=0, c \neq b}^{2^n-1} A_{bc} [\text{Re}(\langle\varphi|b\rangle\langle c|\varphi\rangle) + i \text{Im}(\langle\varphi|b\rangle\langle c|\varphi\rangle)] \\ &= \sum_{b=0}^{2^n-1} A_{bb} |\langle b|\varphi\rangle|^2 + \sum_{b=0}^{2^n-1} \sum_{c=0, c \neq b}^{2^n-1} A_{bc} \left[\frac{1}{2} \left(\left| \frac{\langle b| + \langle c|}{\sqrt{2}} |\varphi\rangle \right|^2 - \left| \frac{\langle b| - \langle c|}{\sqrt{2}} |\varphi\rangle \right|^2 \right) \right. \\ &\quad \left. + \frac{i}{2} \left(\left| \frac{\langle b| - i\langle c|}{\sqrt{2}} |\varphi\rangle \right|^2 - \left| \frac{\langle b| + i\langle c|}{\sqrt{2}} |\varphi\rangle \right|^2 \right) \right] \\ &= \sum_{b=0}^{2^n-1} A_{bb} |\langle b|\varphi\rangle|^2 + \sum_{b=0}^{2^n-1} \sum_{c=0, c \neq b}^{2^n-1} A_{bc} \left[\frac{1}{2} \left(|\langle b|M_{\text{Re}}^{(b,c)}|\varphi\rangle|^2 - |\langle b \oplus 2^{j_0^{(b,c)}}|M_{\text{Re}}^{(b,c)}|\varphi\rangle|^2 \right) \right. \\ &\quad \left. + \frac{i}{2} \left(|\langle b|M_{\text{Im}}^{(b,c)}|\varphi\rangle|^2 - |\langle b \oplus 2^{j_0^{(b,c)}}|M_{\text{Im}}^{(b,c)}|\varphi\rangle|^2 \right) \right] \\ &= \sum_{b=0}^{2^n-1} A_{bb} |\langle b|\varphi\rangle|^2 + \sum_{l \in \mathcal{S}(A) \setminus \{0^n\}} \sum_{b=0}^{2^n-1} \left((-1)^w A_{(b'c')} |\langle b|M_{\text{Re}}^{(l)}|\varphi\rangle|^2 + i(-1)^w A_{[b'c']} |\langle b|M_{\text{Im}}^{(l)}|\varphi\rangle|^2 \right) \\ &= \sum_{l \in \mathcal{S}(A)} \sum_{s \in \{\text{Re}, \text{Im}\}} \sum_{b=0}^{2^n-1} a(l, s, b) p(l, s, b) \end{aligned} \quad (2)$$

Table 1: Definitions of the symbols appearing in Eq. (2).

Symbol	Definition
$\mathcal{S}(A)$	$\{b \oplus c A_{bc} \neq 0\}$
$\oplus$	bit-wise exclusive OR
(b', c', w)	$(b, b \oplus l, 0)$ for $b < b \oplus l$ $(b \oplus 2^{j_0^{(l)}}, b \oplus l \oplus 2^{j_0^{(l)}}, 1)$ for $b > b \oplus l$
$j_0^{(l)}$	$\max\{j l_j = 1\}$
l_j	a value of j th bit of l in the binary expression
$A_{(bc)}$	$(A_{bc} + A_{cb})/2$
$A_{[bc]}$	$(A_{bc} - A_{cb})/2$
$M_{\text{Re}}^{(l)}$	$H(j_0^{(l)}) \left(\prod_{k \in \tilde{\mathcal{T}}^{(l)}} \text{CNOT}(j_0^{(l)}, k) \right)$
$M_{\text{Im}}^{(l)}$	$H(j_0^{(l)}) \left(\prod_{k \in \tilde{\mathcal{T}}^{(l)}} \text{CNOT}(j_0^{(l)}, k) \right) S^\dagger(j_0^{(l)})$
$\tilde{\mathcal{T}}^{(l)}$	$\{j l_j = 1\} \setminus \{j_0^{(l)}\}$
$H(j)$	Hadamard gate acting on the j th qubit
$S(j)$	phase gate acting on the j th qubit
$\text{CNOT}(j, k)$	CNOT gate whose control and target bits are the j th and k th qubits

where $a(l, s, b)$ and $p(l, s, b)$ are respectively defined as

$$a(l, s, b) := \begin{cases} \frac{A_{bb}}{p(l, s)} & l = 0^n \\ \frac{(-1)^w A_{(b'c')}}{p(l, s)} & l \neq 0^n, s = \text{Re} \\ \frac{i(-1)^w A_{[b'c']}}{p(l, s)} & l \neq 0^n, s = \text{Im} \end{cases} \quad (3)$$

and

$$p(l, s, b) := p(l, s) |\langle b | M_s^{(l)} | \varphi \rangle|^2. \quad (4)$$

Note that $M_s^{(0^n)} = I$ and $p(l = 0^n, s) = p(l = 0^n)$ in this representation. $p(l, s)$ is the probability of selecting the measurement operator $M_s^{(l)}$, which can be arbitrarily defined.

Expression of A (Lines 1–2 of Eq. (2))

Unlike existing Pauli-measurement-based methods, Pauli bases are not used to represent A as shown in Eq. (1). This means that the

Pauli-measurement is no longer used to evaluate the expectation value. Instead, the real and imaginary parts of $\langle \varphi | b \rangle \langle c | \varphi \rangle$ are directly measured in the present method. Both the real and imaginary parts of $\langle \varphi | b \rangle \langle c | \varphi \rangle$ can be expressed as the projection of the state $|\varphi\rangle$ on the superposition state of $|b\rangle$ and $|c\rangle$, as shown in the second line of Eq. (2).

Measurement Operators (Lines 2–3 of Eq. (2))

A basis transformation is required to evaluate $|(1/\sqrt{2})(\langle b | \pm \langle c |) | \varphi \rangle|^2$ and $|(1/\sqrt{2})(\langle b | \pm i \langle c |) | \varphi \rangle|^2$ using measurements in the computational basis. Fortunately, this basis transformation can be carried out using simple measurement operators defined as

$$\frac{|b\rangle + |c\rangle}{\sqrt{2}} := (M_{\text{Re}}^{(b,c)})^\dagger |b\rangle \quad (5)$$

$$\frac{|b\rangle + i|c\rangle}{\sqrt{2}} := (M_{\text{Im}}^{(b,c)})^\dagger |b\rangle. \quad (6)$$

Concrete circuit representations of these measurement operators are given in Table 1. A nice

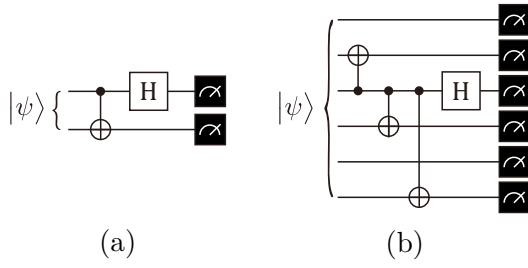


Figure 1: Typical circuits for (a) standard Bell measurements, and (b) extended Bell measurements.

property of these measurement operators is that each of them contains at most $n - 1$ CNOT gates, which is suitable for the NISQ devices. We refer to the present method as the *extended Bell measurements* (XBMs) because the measurement operators $M_{\text{Re}}^{(b,c)}$ and $M_{\text{Im}}^{(b,c)}$ are similar to those for the Bell measurements except for the number of CNOT gates. In Fig. 1, typical circuits for a standard Bell measurements and the extended Bell measurements are shown.

Simultaneous Measurements (Lines 3–4 of Eq. (2)) We proved the following theorem.

Theorem 1 *If $b \oplus c = b' \oplus c'$ and $b' \neq c'$, then $M_{\text{Re}}^{(b,c)} = M_{\text{Re}}^{(b',c')}$ and $M_{\text{Im}}^{(b,c)} = M_{\text{Im}}^{(b',c')}$.*

The proof is given in Appendix A.2. The above theorem states that the number of distinct measurement operators can be drastically reduced. In addition, the superscript $(\bullet)^{(b,c)}$ can be replaced with $(\bullet)^{(b \oplus c)}$ because each distinct measurement operator is identified by $b \oplus c$. Reformulation leads to the fourth line of Eq. (2).

Operator Selecting Model (Lines 4–5 of Eq. (2)) The fourth line of Eq. (2) does not seem to express an expectation value because it does not take the form $\sum_i a_i P(\hat{a} = a_i)$. Here, we introduce an operator selecting model $p(l, s)$ that represents the probability of selecting the measurement operator $M_s^{(l)}$, where $l \in \mathcal{S}(A)$ and $s \in \{\text{Re}, \text{Im}\}$. Here, $\mathcal{S}(A)$ is a set of $b \oplus c$ where A_{bc} has non-zero components, as defined in Table 1. Introducing $p(l, s)$, $\langle \varphi | A | \varphi \rangle$ is eventually represented as on the rightmost side of Eq. (2) and takes the form $\sum_i a_i P(\hat{a} = a_i)$. Although the concrete description of $p(l, s)$ is arbitrary the appropriate definition of $p(l, s)$ reduces the variance of the expectation estimation.

When we regard Eq. (2) as expressing the expectation value in a stochastic process, an event is selecting a measurement operator $M_s^{(l)}$ with probability $p(l, s)$ followed by measuring a bit-string b of a state $M_s^{(l)}|\varphi\rangle$ in the computational basis, and then obtaining a corresponding outcome. Denoting a complex random variable corresponding to such an event as $\hat{a}$, we have

$$\mathbb{E}_{l,s,b \sim p(l,s,b)}[\hat{a}] = \langle \varphi | A | \varphi \rangle. \quad (7)$$

In the following, we simply express $\mathbb{E}_{l,s,b \sim p(l,s,b)}[\hat{a}]$ as $\mathbb{E}_p[\hat{a}]$.

Significance of Eq. (2) There are three important aspects to the expression in Eq. (2).

1. There are fewer distinct measurement operators.
2. Each measurement operator has a simple explicit expression.
3. There are fewer CNOT gates in each measurement operator.

Equation (2) contains only $|\mathcal{S}(A) \times \{\text{Re}, \text{Im}\}|$ measurement operators, I , $\{M_{\text{Re}}^{(l)}\}_{l \in \mathcal{S}(A) \setminus \{0^n\}}$, and $\{M_{\text{Im}}^{(l)}\}_{l \in \mathcal{S}(A) \setminus \{0^n\}}$. The number of distinct measurement operators depends on the structure of A , which is discussed further in Sec. 3. A simple explicit expression of each measurement operator is beneficial from the viewpoint of time complexity for circuit construction. The number of CNOT gates included in each measurement operator is at most $n - 1$, which is suitable for the NISQ device because most of the real-device error comes from two-qubit entangling gates. In summary, qualitatively speaking, the construction of measurement operators is fast and the evaluation error on a real quantum computer using these operators is expected to be small. Further details of the complexity analyses are given in Sec. 3.

2.3 Upper Bound of Variance

An upper bound of the variance of $\hat{a}$ associated with the XBM can be easily calculated as

$$\begin{aligned} \text{Var}_p[\hat{a}] &= \mathbb{E}_p[|\hat{a}|^2] - |\mathbb{E}_p[\hat{a}]|^2 \\ &\leq \mathbb{E}_p[|\hat{a}|^2] \\ &= \sum_{l,s,b} |a(l, s, b)|^2 p(l, s, b), \end{aligned} \quad (8)$$

which depends on the operator selecting model $p(l, s)$. A simple choice of $p(l, s)$ is a uniform distribution over l and s whereas a more sophisticated choice can reduce the variance of $\hat{a}$ (see Appendix B for details).

Here, we assume that $|A|_{\max}^2 = O(\text{tr}(A^2)/q)$ where $|A|_{\max} := \max_{b,c} \{|A_{bc}|\}$ is the maximum norm of A and q the number of non-zero components of A , respectively. Under this assumption, no matter which operator selecting model is used, the upper bound of the variance of $\hat{a}$ can be roughly estimated as

$$\text{Var}_p[\hat{a}] \leq m^2 |A|_{\max}^2 = O\left(\frac{m^2}{q} \text{tr}(A^2)\right) \quad (9)$$

where

$$m := |\mathcal{S}(A) \times \{\text{Re}, \text{Im}\}| \quad (10)$$

is the number of distinct measurement operators (see Appendix C.2 for details).

2.4 Relationship with the Classical Shadow

Our method is closely related to the classical shadow with random Clifford measurements [31]. The main differences are the unitary ensemble and the operator selecting model. The random Clifford circuits are used in the classical shadow whereas m distinct measurement operators are used in the XBM. In addition, only the uniform operator selecting model is considered in the original classical shadow whereas an arbitrary operator selecting model is considered in the XBM.

Additionally, the number of two-qubit entangling gates in each measurement operator is different. The classical shadow with random Clifford measurements requires $O(n^2/\log(n))$ two-qubit entangling gates for each measurement operator whereas the XBM requires at most $n - 1$ CNOT gates for each.

The upper bound of the variance associated with the classical shadow with the random Clifford measurements is $O(\text{tr}(A^2))$ [31]. Hence, when $m^2 \ll q$, our method has an upper bound tighter than that of the classical shadow. An example of this case is that non-zero components of A are densely filled in the tight bandwidth. When the bandwidth of A is $k = O(n^c)$, and the non-zero components are densely filled in the

band,

$$\text{Var}_p[\hat{a}] = O\left(\frac{n^{c+1}}{2^n} \text{tr}(A^2)\right) \quad (11)$$

(see Appendix C.2 for other cases). Such a matrix frequently appear in the case of the FEM as mentioned in the introduction. In this case, the adoption of our method exponentially improves the variance relative to the classical shadow.

The relationship between the XBM and classical shadow is further discussed in Appendix D, which gives the explicit form of the classical shadow associated with the XBM method.

3 Algorithm and Complexity

The algorithms for term grouping and calculating an expectation value using the XBM method are presented as Algorithms 1 and 2, respectively. These algorithms follow the equations described in Eq. (2) (see Appendix A for details). The greatest advantage of the proposed method, the XBM method, is that the number of unique circuits required to estimate $\langle \psi_0 | A | \psi_1 \rangle$ can be rapidly reduced. In this section, the reduction ratio of the number of unique circuits and the computational cost of the XBM method are discussed.

3.1 Number of Distinct Operators

In this paper, an s -sparse matrix and band matrix with bandwidth k are defined as follows.

Definition 1 (s -sparse matrix A [35])

$A \in \mathbb{C}^{2^n \times 2^n}$ is s -sparse if each column has at most s nonzero entries.

Definition 2 (band matrix A) $A \in \mathbb{C}^{2^n \times 2^n}$ is a band matrix with bandwidth k if $A_{ij} = 0$ for all i, j such that $|i - j| > k$.

According to these definitions, the following theorems hold.

Theorem 2 (s -sparse matrix A) Let $A \in \mathbb{C}^{2^n \times 2^n}$ and $|\psi_0\rangle, |\psi_1\rangle \in \mathbb{C}^{2^n}$. When A is an s -sparse matrix, there exists an algorithm such that the number of quantum circuits required to evaluate $\langle \psi_0 | A | \psi_1 \rangle$ is 2^{n+1} in the worst case.

Algorithm 1: Algorithm for term grouping in the XBM method

```
input    : Non-zero matrix elements and corresponding indices,  $\{b, c, A_{bc} | A_{bc} \neq 0\}$ 
output  : Measurement operators  $M_s^{(l)}$  and outcomes  $\alpha(l, s, b)$ 
initialize:  $M_s^{(l)} \leftarrow \text{None}$  and  $\alpha(l, s, b) \leftarrow 0$  for all  $l, s, b$ 
for  $(b, c) \in \{(b, c) | A_{bc} \neq 0\}$  do
     $l \leftarrow b \oplus c$ 
    if  $l \neq 0^n$  then
         $j_0^{(l)} \leftarrow \max\{j | l_j = 1\}$ 
    end
    /* Compute and store measurement operators */
    if Measurement operators  $M_{\text{Re}}^{(l)}$  and  $M_{\text{Im}}^{(l)}$  have not yet been computed then
        if  $l = 0^n$  then
             $M_{\text{Re}}^{(l)} \leftarrow I$ 
        else
             $M_{\text{Re}}^{(l)}, M_{\text{Im}}^{(l)} \leftarrow (\text{see Table 1})$ 
        end
    end
    /* Compute and store outcomes */
    if  $b = c$  then
         $\alpha(l, \text{Re}, b) \leftarrow \alpha(l, \text{Re}, b) + A_{bc}$ 
    else if  $b < c$  then
         $\bar{b} \leftarrow b \oplus 2^{j_0^{(l)}}$ 
         $\alpha(l, \text{Re}, b) \leftarrow \alpha(l, \text{Re}, b) + A_{bc}/2$ 
         $\alpha(l, \text{Re}, \bar{b}) \leftarrow \alpha(l, \text{Re}, \bar{b}) - A_{bc}/2$ 
         $\alpha(l, \text{Im}, b) \leftarrow \alpha(l, \text{Im}, b) + iA_{bc}/2$ 
         $\alpha(l, \text{Im}, \bar{b}) \leftarrow \alpha(l, \text{Im}, \bar{b}) - iA_{bc}/2$ 
    else if  $b > c$  then
         $\bar{c} \leftarrow c \oplus 2^{j_0^{(l)}}$ 
         $\alpha(l, \text{Re}, c) \leftarrow \alpha(l, \text{Re}, c) + A_{bc}/2$ 
         $\alpha(l, \text{Re}, \bar{c}) \leftarrow \alpha(l, \text{Re}, \bar{c}) - A_{bc}/2$ 
         $\alpha(l, \text{Im}, c) \leftarrow \alpha(l, \text{Im}, c) - iA_{bc}/2$ 
         $\alpha(l, \text{Im}, \bar{c}) \leftarrow \alpha(l, \text{Im}, \bar{c}) + iA_{bc}/2$ 
    end
end
```

Theorem 3 (band matrix A) Let $A \in \mathbb{C}^{2^n \times 2^n}$ and $|\psi_0\rangle, |\psi_1\rangle \in \mathbb{C}^{2^n}$. When the bandwidth of A is $k > 0$, there exists an algorithm such that the number of quantum circuits required to evaluate $\langle \psi_0 | A | \psi_1 \rangle$ is $2((n-r)k + 2^r)$ in the worst case where $r = \lceil \log_2 k \rceil$.

Here, we define the upper bound of the number of the unique circuits for an n -qubit system with bandwidth k as

$$\overline{m}(n, k) := \begin{cases} 2((n-r)k + 2^r) & k > 0 \\ 1 & k = 0 \end{cases} \quad (12)$$

where $r := \lceil \log_2 k \rceil$. Note that when evaluating $\langle \varphi | A | \varphi \rangle$ whose A has non-zero diagonal components, the upper bound can be reduced to $2((n-r)k + 2^r) - 1$ for $k > 1$ (see Appendix E.2).

The proofs are given in Appendix E. Note that a fully dense matrix can be viewed as a 2^n -sparse matrix or a band matrix with bandwidth $k = 2^n - 1$.

Unfortunately, we find that the evaluation of the number of circuits in Algorithm 1 is tight; i.e., there exists an s -sparse matrix that requires 2^{n+1} unique circuits, which is the same as the case for the fully dense matrix. However, if A is a band matrix, then at least the number of

Algorithm 2: Algorithm for calculating an expectation value in the XBM method. Note that median-of-means or others can be used instead of the arithmetic mean in the last line.

```

input   : Number of shots  $N$ , operator selecting model  $p(l, s)$ , measurement operators
             $\{M_s^{(l)}\}$ , quantum state  $|\varphi\rangle$ , outcomes  $\{\alpha(l, s, b)\}$ 
output  : Expectation value  $x$ 
for  $i = 1$  to  $N$  do
    /* Pick a measurement operator                                     */
     $l, s \sim p(l, s)$  ;                                     // Classical computing
     $M \leftarrow M_s^{(l)}$  ;                               // Access to the classical memory
    /* Measure a bit string and update  $x$                              */
     $b \sim |\langle b|M|\varphi\rangle|^2$  ;                             // Quantum computing
     $x_i \leftarrow \alpha(l, s, b)$  ;           // Access to the classical memory and classical computing
end
 $x \leftarrow \sum_i x_i / N$  ;                               // Take arithmetic mean

```

unique circuits required to estimate $\langle\psi_0|A|\psi_1\rangle$ can be reduced to $2((n-r)k + 2^r)$, where k is the bandwidth and $r = \lceil \log_2 k \rceil$. When A is sufficiently tight (i.e., bandwidth $k = O(n^c)$), $2((n-r)k + 2^r) = O(n^{c+1})$. Note that the number of unique Pauli strings required to evaluate $\langle\psi_0|A|\psi_1\rangle$ using naïve Pauli measurements is $2^{n+1}((n-r)k + 2^r)$ (see Sec. E.4). The XBM thus outperforms the naïve Pauli measurements in terms of the number of unique circuits required to evaluate $\langle\psi_0|A|\psi_1\rangle$. This exponential improvement is due to the difference in the basis of A , but the number of unique circuits when using the XBM method does not appreciably increase even if A is represented in the Pauli basis (see Sec. E.5).

As mentioned before, our method is the generalization of [7]. In their formulation, both $\langle\varphi|A^2|\varphi\rangle = \langle\varphi|(B + C^{(0)} + C^{(1)})|\varphi\rangle$ and $\langle b|A|\varphi\rangle$ are calculated to solve the one-dimensional Poisson equation, and the numbers of unique circuits required to evaluate $\langle\varphi|A^2|\varphi\rangle = \langle\varphi|(B + C^{(0)} + C^{(1)})|\varphi\rangle$ and $\langle b|A|\varphi\rangle$ are $4n+1$ and $2n+1$, respectively, where the bandwidth of B , $C^{(0)}$, $C^{(1)}$ and A are 2, 0, 0 and 1, respectively. This is consistent with $\overline{m}(n, k=2) + \overline{m}(n, k=0) + \overline{m}(n, k=0) = 4n+1+1 = 4n+2$ and $\overline{m}(n, k=1) = 2n+2$ (see Eq. (12)), where the +1 difference comes from the counting of an unnecessary term, $\text{Im}(\langle\varphi|b\rangle\langle b|\varphi\rangle) = 0$.

3.2 Time Complexity for Term Grouping

We assume that data $\{(b, c, A_{bc}) | A_{bc} \neq 0\}$ are stored in the memory of the classical computer,

where A_{bc} is the (b, c) component of the matrix A . For each $(b, c) \in \{(b, c) | A_{bc} \neq 0\} := \mathcal{B}$, bit strings to be measured are calculated from b and $j_0^{(b \oplus c)} = \max\{j | (b \oplus c)_j = 1\}$ for $b \neq c$. The computational cost of calculating $j_0^{(b \oplus c)}$ with classical computers is $O(n)$, and the total time complexity for term grouping is thus $O(nd)$, where d is the size of $\mathcal{B}$, which is the number of non-zero elements of A .

Note that in Algorithm 1, measurement operators $M_s^{(l)}$ are simultaneously calculated with term grouping. Each measurement circuit can be built in $O(n)$ time (see definition of $M_s^{(l)}$ in Table 1), and its calculation is only carried out when $M_s^{(l)}$ is not yet calculated. Therefore, the total time complexity of building measurement circuits is $O(nm)$ where $m \leq d$ is the number of distinct $b \oplus c$ values among (b, c) pairs with non-zero A_{bc} values, which does not affect the total time complexity of term grouping.

3.3 Gate Counts for Measurement Circuits

As described in Table 1, the maximum number of gate counts is $n+1$ for each unique circuit; i.e., there are $n-1$ CNOT gates, one Hadamard gate and one phase gate. Hence, the gate count for each measurement circuit is $O(n)$.

4 Experiments

We implement our model with Qiskit version 0.25.0 [34]. All experiments are carried out on a single Intel(R) Core(TM) i9-9900X CPU @ 3.50GHz and ibm_kawasaki of an IBM-Q

system. To compare with the existing methods, QWC (qubit-wise commuting) [21, 27, 28] and GC (general commuting) [24] are conducted with the same machine. The Python function, `group_into_tensor_product_basis_sets`, implemented in OpenFermion [36], is used for QWC whereas Python code provided by one of the authors of [24] is used for GC. In GC, the Bron-Kerbosch algorithm [37] is used to solve the minimum clique covering problem.

In our method, the XBM method, $\text{Re}(\langle\varphi|b\rangle\langle c|\varphi\rangle)$ and $\text{Im}(\langle\varphi|b\rangle\langle c|\varphi\rangle)$ can be separately evaluated. To take this advantage, we consider the two following cases in which $\text{Im}(\langle\varphi|b\rangle\langle c|\varphi\rangle)$ does not need to be estimated.

1. $A \in \mathbb{R}^{2^n \times 2^n}$ and only $\text{Re}(\langle\psi_0|A|\psi_1\rangle)$ is required. For example, A is real symmetric matrix and $|\psi_0\rangle = |\psi_1\rangle$.

2. $\exists\phi_0, \phi_1$ s.t. $e^{i\phi_0}|\psi_0\rangle, e^{i\phi_1}|\psi_1\rangle \in \mathbb{R}^{2^n}$

In these cases, only $\text{Re}(\langle\varphi|b\rangle\langle c|\varphi\rangle)$ needs to be estimated because

$$\begin{aligned} \text{Re}(\langle\varphi|A|\varphi\rangle) &= \sum_{b,c} \text{Re}(A_{bc}) \text{Re}(\langle\varphi|b\rangle\langle c|\varphi\rangle) \\ &\quad - \sum_{b,c} \text{Im}(A_{bc}) \text{Im}(\langle\varphi|b\rangle\langle c|\varphi\rangle) \end{aligned}$$

and

$$\begin{aligned} \text{Im}(\langle\varphi|A|\varphi\rangle) &= \sum_{b,c} \text{Im}(A_{bc}) \text{Re}(\langle\varphi|b\rangle\langle c|\varphi\rangle) \\ &\quad + \sum_{b,c} \text{Re}(A_{bc}) \text{Im}(\langle\varphi|b\rangle\langle c|\varphi\rangle). \end{aligned}$$

The computational cost is reduced by half in these cases and we thus refer these to these cases by “XBM (half)” in the following. Note that in the case that $A \in \mathbb{C}^{2^n \times 2^n}$ and $|\psi_0\rangle, |\psi_1\rangle \in \mathbb{C}^{2^n}$, even if $|\psi_0\rangle = |\psi_1\rangle$ and A is an Hermitian matrix, which is the case that $\langle\psi_0|A|\psi_1\rangle \in \mathbb{R}$, both $\text{Re}(\langle\varphi|b\rangle\langle c|\varphi\rangle)$ and $\text{Im}(\langle\varphi|b\rangle\langle c|\varphi\rangle)$ have to be evaluated.

$2|0\rangle\langle 1| \otimes A$ (Eq. (S.2)) is used instead of A when evaluating the case that $|\psi_0\rangle \neq |\psi_1\rangle$. In QWC and GC, the matrix A is converted to the sum of the Pauli bases before adopting the term grouping algorithms.

4.1 Term Grouping

4.1.1 Setups

We first show that the present method is more efficient in terms of both the number of unique cir-

cuits and the computational cost for term grouping than existing methods. We conduct the following three experiments:

1. We compare the number of unique circuits and time required for term grouping between different methods when $|\psi_0\rangle = |\psi_1\rangle$.
2. We repeat the first experiment but for $|\psi_0\rangle \neq |\psi_1\rangle$.
3. We estimate the number of unique circuits for the XBM method when the non-zero elements of A are randomly placed.

In the first and second experiments, all components of $A \in \mathbb{C}^{2^n \times 2^n}$ are randomly set to non-zero values. A is then restricted to be a band matrix by setting $A_{bc} = 0$ for $|b - c| > k$, where $k = 0, 1, \dots, 2^{n-1}$. In the third experiment, we observe the term groupings while changing the number of non-zero elements, d , that are randomly placed in A . The number of unique circuits is obtained 10 times for each value of d . The mean and standard deviation are then calculated.

4.1.2 Results

Experiment 1 Figure 2(a) shows the dependency of the number of unique circuits on n , in the case that $|\psi_0\rangle = |\psi_1\rangle$ and the bandwidth $k = 3$. The XBM and GC results have similar tendencies that are much better than the tendencies of Naïve and QWC in terms of the reduction ratio of the unique circuits. When the imaginary part of $\langle\varphi|b\rangle\langle c|\varphi\rangle$ is not required, as in the case of XBM (half), the number of unique circuits is lower for the XBM method than for the GC method.

Figure 2(b) shows the dependency of the computational time for term grouping on n when the bandwidth $k = 3$. Obviously, the XBM method is much faster than all other methods for large n . This is because GC and QWC methods heuristically solve a minimum clique cover problem that is NP-hard, whereas the XBM method analytically groups the terms as shown in Sec. 2. Note that this qualitative tendency of the comparison results does not change for any bandwidth as shown in Fig. 3(a) and Fig. 3(b).

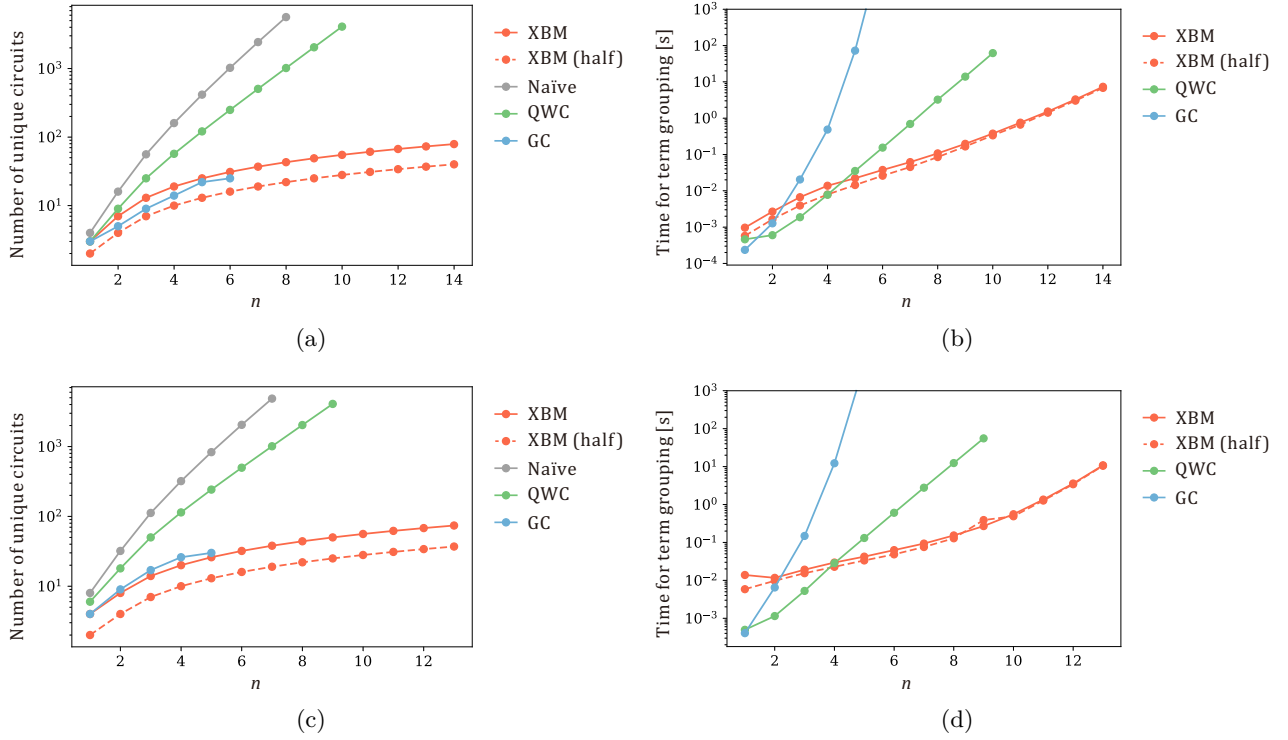


Figure 2: Number of unique circuits in estimating $\langle \psi_0 | A | \psi_1 \rangle$ and the time required for term grouping when the bandwidth of A is $k = 3$: (a)(b) $|\psi_0\rangle = |\psi_1\rangle$; (c)(d) $|\psi_0\rangle \neq |\psi_1\rangle$. XBM: present method, XBM (half): present method but only evaluating the real part of $\langle \varphi | b \rangle \langle c | \varphi \rangle$, Naïve: Pauli measurements without term grouping, QWC: Pauli measurements with grouping associated with the qubit-wise commutativity [21, 27, 28], GC: Pauli measurements with grouping associated with the general commutativity [24].

Experiment 2 Figures 2(c) and 2(d) show the results for the case that $|\psi_0\rangle \neq |\psi_1\rangle$. In this case, regardless of whether the imaginary part of $\langle \varphi | b \rangle \langle c | \varphi \rangle$ is required, the number of unique circuits is lower for the XBM method than for all other methods. This is because the number of unique circuits for $|0\rangle\langle 1| \otimes 2A$ is the same as that for A for the XBM method, but not for the other methods. The tendency of the time required for term grouping against n in the case that $|\psi_0\rangle \neq |\psi_1\rangle$ is the same as that in the case that $|\psi_0\rangle = |\psi_1\rangle$.

Experiment 3 The third experiment is carried out to show that the XBM method is more efficient in terms of the number of unique circuits when A is a band-matrix. Figure 4 shows the results of the number of unique circuits for the XBM method when the non-zero elements of A are randomly placed; this case is referred to as *random-A*. The upper bound of the number of unique circuits when A is a band-matrix with a bandwidth $k = 3$ is also shown in the figure (blue line); this case is referred to as *band-A*.

The number of unique circuits required to evaluate *random-A* is analytically derived as follows. Let A have d non-zero elements at random. Here, we consider a box that contains a total of N^2 balls of N types, where $N = 2^n$. In fact, the number of unique circuits is twice the expectation number of ball types when one samples d balls from the box. The probability that the balls are of type k when d balls are sampled from a box containing a total of N^2 balls of N types, $P(k; d, N)$, is

$$P(k; d, N) = \frac{\mathfrak{N}(k, d, N)}{N^2 C_d}$$

where

$$\mathfrak{N}(k, d, N) = N C_k \sum_{i=1}^{k-1} \frac{N C_i}{N C_i} \mathfrak{N}(i, d, N).$$

The expectation value is then given as

$$E(d, N) = \sum_{k=1}^d k P(k; d, N).$$

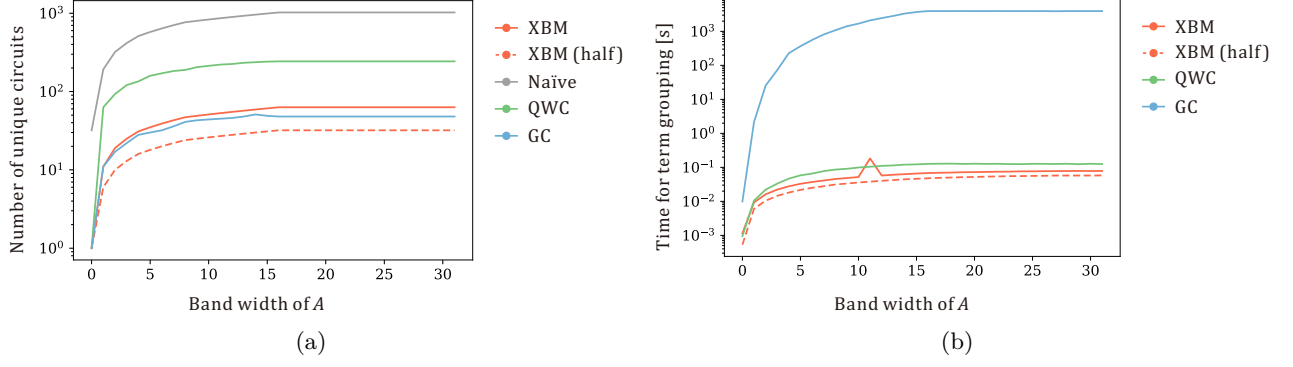


Figure 3: Number of unique circuits in estimating $\langle \psi_0 | A | \psi_0 \rangle$ and the time required for term grouping with respect to the bandwidth of A , for a five-qubit system. XBM: present method, XBM (half): present method but only evaluating the real part of $\langle \varphi | b \rangle \langle c | \varphi \rangle$, Naïve: Pauli measurements without term grouping, QWC: Pauli measurements with grouping associated with the qubit-wise commutativity [21, 27, 28], GC: Pauli measurements with grouping associated with the general commutativity [24].

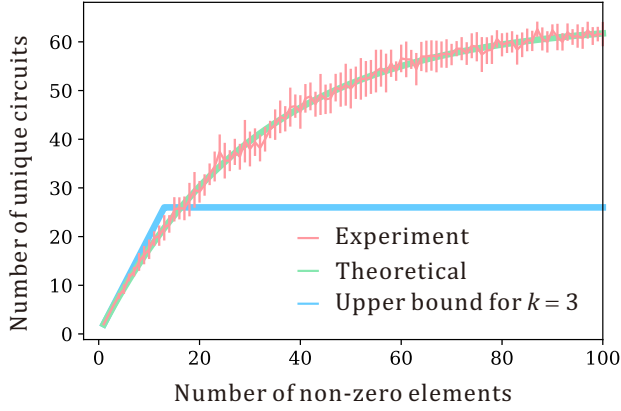


Figure 4: Results when the non-zero elements of A are randomly placed for the five-qubit system. Red and green lines are the experimental and analytic results of the XBM, respectively. The blue line is the upper bound when all non-zero elements are placed in a $k=3$ band, $\min(d, \overline{m}(5, 3)) = \min(d, 2((5 - \lceil \log_2 3 \rceil) \cdot 3 + 2^{\lceil \log_2 3 \rceil}))$

The green line in Fig. 4 denotes the expectation value $E(d, N = 2^5)$ and is consistent with the red line, which is the mean of the experimental results. The figure shows that the number of unique circuits for *random-A* is much larger than that for *band-A* when the number of nonzero elements of A is sufficiently large. For example, when $d = 50$, the expectation number of unique circuits for *random-A* is approximately 52. Meanwhile, the number of unique circuits is at most 26 when all 50 non-zero elements are put into a bandwidth $k = 3$. This indicates that adoption of the XBM method can reduce the number of unique circuits more efficiently when the bandwidth of A is narrower.

4.2 Real-device Experiments

4.2.1 Setups

In real-device experiments, we use the uniform operator selecting model $p(l, s) = 1/m$. For simplicity, all measurement operators are used at once instead of sampling l and s . We set the number of shots for each measurement operator at 8192, which is the maximum number for the IBM-Q system. We use the `ibm_kawasaki` backend for all real-device experiments because its CNOT error and readout error are smaller than those of other backends. Both the elapsed wall time for preprocessing and execution time on the IBM-Q system are recorded. Here, preprocessing includes term grouping, calculating the coefficients of Pauli bases, building measurement circuits, and all other calculations conducted on a classical computer before carrying out quantum computing.

We conduct four real device experiments.

1. We compare the relative errors of $\langle \psi_0 | A | \psi_1 \rangle$ and execution time for each method when $|\psi_0\rangle = |\psi_1\rangle$.
2. We evaluate the relative errors of $\langle \varphi | A_{bc} | \varphi \rangle$ where A_{bc} has only one non-zero element in the b th row and c th column.
3. We compare the relative errors of $\langle \psi_0 | A | \psi_1 \rangle$ and execution time for each method when A is a band matrix and $|\psi_0\rangle = |\psi_1\rangle$.
4. We repeat the first experiment, but $|\psi_0\rangle \neq |\psi_1\rangle$.

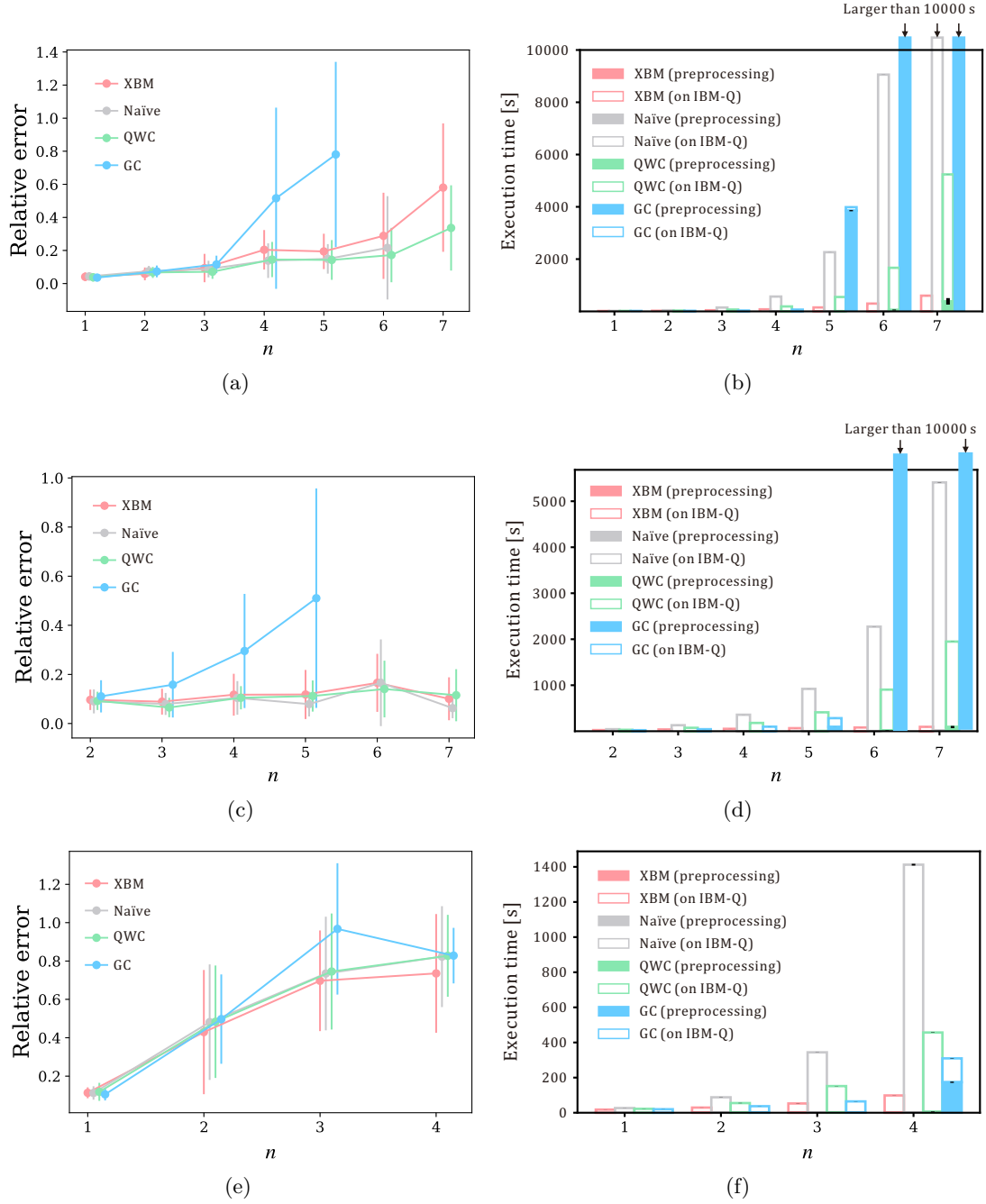


Figure 5: Results of the real device evaluation of $\langle \psi_0 | A | \psi_1 \rangle$. (a)(b) $|\psi_0\rangle = |\psi_1\rangle$; (c)(d) $|\psi_0\rangle = |\psi_1\rangle$ and $k = 3$; (e)(f) $|\psi_0\rangle \neq |\psi_1\rangle$. XBM: present method, Naïve: Pauli measurements without term grouping, QWC: Pauli measurements with grouping associated with the qubit-wise commutativity [21, 27, 28], GC: Pauli measurements with grouping associated with the general commutativity [24]. *method* (preprocessing) and *method* (on IBM-Q) denote the execution time of classical preprocessing required before quantum processing and of quantum processing on the IBM-Q system, respectively.

Here, we define the relative errors as

$$\frac{|\langle \psi_0 | A | \psi_1 \rangle_{\text{real device}} - \langle \psi_0 | A | \psi_1 \rangle_{\text{GT}}|}{|\langle \psi_0 | A | \psi_1 \rangle_{\text{GT}}|}$$

where $(\bullet)_{\text{real device}}$ and $(\bullet)_{\text{GT}}$ are the evaluation

value of $(\bullet)$ on the real device and the ground truth of $(\bullet)$, respectively.

In the first, third and fourth experiments, all the components of A are randomly set as $\text{Re}(A_{bc}), \text{Im}(A_{bc}) \in [-100, 100]$. In preparing $|\psi_0\rangle$ and $|\psi_1\rangle$, R_x , R_y and R_z gates are randomly

set for each qubit. The experiments are carried out 10 times for each condition, and the mean and standard deviation of the results are calculated.

In the second experiment, only one component, A_{bc} , is set to be non-zero (i.e., $\text{Re}(A_{bc}), \text{Im}(A_{bc}) \in [-100, 100]$), whereas all other components are set to zero. Unlike the case in the other experiments, $|\varphi\rangle$ is prepared by acting on all qubits with Hadamard gates to mitigate the effect of the randomness of the $|\varphi\rangle$ preparation. This experiment is repeated for $b = 0, 1, \dots, 2^n - 1$ and $c = 0, 1, \dots, 2^n - 1$. As a result, the relative errors for all components of A are obtained. Only XBM and Naïve experiments are carried out here. The experiments are carried out 10 times for each (b, c) , and the mean results are calculated.

4.2.2 Results

Experiment 1 As shown in Fig. 5(a), the relative errors of the XBM method are much better than those of the GC method but worse than those of the Naïve and QWC methods for $n \geq 4$ when $|\psi_0\rangle = |\psi_1\rangle$. This trend can be explained by the circuit complexity. The XBM method uses multiple CNOT gates whereas the Naïve and QWC methods do not, and the circuit for the GC method is much more complex than the circuits of the other methods. Although the XBM method is slightly inferior in terms of the relative error compared with the Naïve and QWC methods, the total execution time for the XBM method is much shorter than the times for the other methods (Fig. 5(b)). The execution time on a real device for the XBM method is slightly longer than that for the GC method, which depends on the number of unique circuits. However, the total execution time for the XBM method is much shorter than that for the GC method because the XBM method performs term grouping much more rapidly than the GC method. In the case of the XBM method, the execution time for term grouping is negligible compared with the execution time on the IBM-Q system. In summary, the proposed method, namely the XBM method, can evaluate $\langle\psi_0|A|\psi_0\rangle$ with slightly worse accuracy than the Naïve and QWC methods but in much less time.

Experiment 2 Figure 6 compares the element-wise relative error between the XBM and Naïve

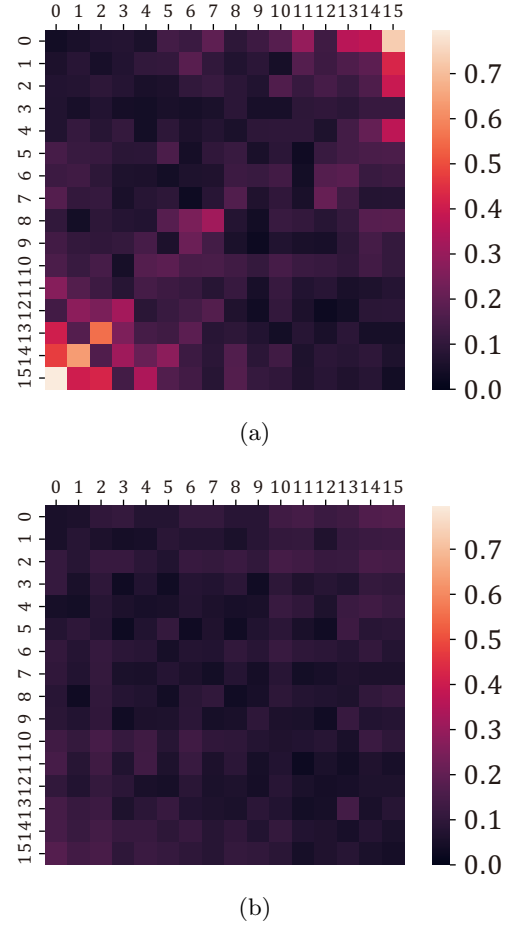


Figure 6: Element-wise relative errors of $\langle\varphi|A_{rc}|\varphi\rangle$ for a four-qubit system evaluated on a real device: (a) XBM method; (b) Naïve method.

methods. In the case of the XBM method, the relative errors are larger for the anti-diagonal components than for the other components, whereas all components have similar relative errors in the case of the Naïve method. This result is consistent with that of the XBM method using the most CNOT gates in the evaluation of anti-diagonal components. Note that the number of CNOT gates required for the XBM method to evaluate the (b, c) component of A is

$$\max\left(0, \sum_{i=0}^{2^n-1} b_i \oplus c_i - 1\right).$$

This indicates that the XBM method can evaluate $\langle\psi_0|A|\psi_0\rangle$ with much smaller error when A is a band matrix.

Experiment 3 Figure 5(c) shows that the relative errors of the XBM, Naïve and QWC methods are small in the case that A is a band matrix; the

errors approximately range $0.1 - 0.15$ and they are all comparable for at least $n \leq 7$. Meanwhile, the XBM method can evaluate $\langle \psi_0 | A | \psi_0 \rangle$ much more rapidly than the Naïve and QWC methods (Fig. 5(d)). It is thus concluded that the XBM method can evaluate $\langle \psi_0 | A | \psi_0 \rangle$ with the same accuracy as the Naïve and QWC methods and in a much shorter time when A is a band matrix.

Experiment 4 Figures 5(e) and 5(f) show the results for the case that $|\psi_0\rangle \neq |\psi_1\rangle$. As shown in Fig. 5(e), although the XBM method can evaluate $\langle \psi_0 | A | \psi_1 \rangle$ much more rapidly than all the other methods (Fig. 5(f)), the relative errors are much larger than in the case that $|\psi_0\rangle = |\psi_1\rangle$ for all methods. This is because in the case that $|\psi_0\rangle \neq |\psi_1\rangle$, $|\psi_0, \psi_1\rangle = (|0\rangle|\psi_0\rangle + |1\rangle|\psi_1\rangle)/\sqrt{2}$ is used to evaluate $\langle \psi_0 | A | \psi_1 \rangle$, which requires a large number of two-qubit entangling gates as shown in Fig. S.1. As pointed out above, the evaluation error increases with the number of two-qubit entangling gates, such as CNOT gates. Hence, to suppress the error in the case that $|\psi_0\rangle \neq |\psi_1\rangle$, reducing the two-qubit entangling gate error or adopting an alternative approach is required.

5 Conclusions

We proposed a method of evaluating $\langle \psi_0 | A | \psi_1 \rangle$ where $A \in \mathbb{C}^{2^n \times 2^n}$ and $|\psi_0\rangle, |\psi_1\rangle \in \mathbb{C}^{2^n}$. Using our method of XBMs, the number of unique circuits required to estimate the expectation value can be reduced to $2(2^r + k(n - r))$ in the worst case, where k is the bandwidth of A and $r := \lceil \log_2 k \rceil$. The greatest advantage of the XBM method over existing methods is the computational time required to prepare the simultaneous measurements. The time complexity of the XBM method including term grouping and building the measurement circuits is $O(nd)$, where d is the number of non-zero elements of A and the gate counts for the measurement operators are $O(n)$ for each unique circuit. In addition, when non-zero elements of A are densely filled in the bandwidth k , the variance of the expectation evaluation is $O((n^{c+1}/2^n)\text{tr}(A^2))$, which is smaller than that of existing methods. The real device experiments show that although there is a trade-off between accuracy and speed, the XBM method has an appreciable advantage over existing meth-

ods in terms of the computational cost, especially when A is a band matrix.

6 Acknowledgement

This work is partly supported by UTokyo Quantum Initiative. We thank Prof. Imoto for insightful comments.

References

- [1] Peruzzo, A. et al. A variational eigenvalue solver on a photonic quantum processor. *Nature communications*, 5(1):1–7, 2014. DOI: [10.1038/ncomms5213](https://doi.org/10.1038/ncomms5213).
- [2] Cerezo, M. et al. Variational quantum algorithms. *Nature Reviews Physics*, 3(9):625–644, 2021. DOI: [10.1038/s42254-021-00348-9](https://doi.org/10.1038/s42254-021-00348-9).
- [3] Preskill, J. Quantum computing in the NISQ era and beyond. *Quantum*, 2:79, 2018. DOI: [10.22331/q-2018-08-06-79](https://doi.org/10.22331/q-2018-08-06-79).
- [4] Huang, H.Y., Bharti, K. and Rebentrost, P. Near-term quantum algorithms for linear systems of equations with regression loss functions. *New Journal of Physics*, 23(11):113021, nov 2021. DOI: [10.1088/1367-2630/ac325f](https://doi.org/10.1088/1367-2630/ac325f).
- [5] Bravo-Prieto, C., LaRose, R., Cerezo, M., Subasi, Y., Cincio, L. and Coles, P.J. Variational quantum linear solver. *arXiv preprint arXiv:1909.05820*, 2019. DOI: [10.48550/arXiv.1909.05820](https://doi.org/10.48550/arXiv.1909.05820).
- [6] Xu, X., Sun, J., Endo, S., Li, Y., Benjamin, S.C. and Yuan, X. Variational algorithms for linear algebra. *Science Bulletin*, 66(21):2181–2188, 2021. DOI: [10.1016/j.scib.2021.06.023](https://doi.org/10.1016/j.scib.2021.06.023).
- [7] Liu, H.L. et al. Variational quantum algorithm for the poisson equation. *Physical Review A*, 104(2):022418, 2021. DOI: [10.1103/physreva.104.022418](https://doi.org/10.1103/physreva.104.022418).
- [8] Sato, Y., Kondo, R., Koide, S., Takamatsu, H. and Imoto, N. Variational quantum algorithm based on the minimum potential energy for solving the poisson equation. *Physical Review A*, 104(5):052409, 2021. DOI: [10.1103/physreva.104.052409](https://doi.org/10.1103/physreva.104.052409).
- [9] Farhi, E., Goldstone, J. and Gutmann, S. A quantum approximate optimization al-

- gorithm. *arXiv preprint arXiv:1411.4028*, 2014. DOI: [10.48550/arXiv.1411.4028](https://doi.org/10.48550/arXiv.1411.4028).
- [10] Zhou, L., Wang, S.T., Choi, S., Pichler, H. and Lukin, M.D. Quantum approximate optimization algorithm: Performance, mechanism, and implementation on near-term devices. *Physical Review X*, 10(2):021067, 2020. DOI: [10.1103/PhysRevX.10.021067](https://doi.org/10.1103/PhysRevX.10.021067).
- [11] Harrigan, M.P. et al. Quantum approximate optimization of non-planar graph problems on a planar superconducting processor. *Nature Physics*, 17(3):332–336, 2021. DOI: [10.1038/s41567-020-01105-y](https://doi.org/10.1038/s41567-020-01105-y).
- [12] Romero, J., Olson, J.P. and Aspuru-Guzik, A. Quantum autoencoders for efficient compression of quantum data. *Quantum Science and Technology*, 2(4):045001, 2017. DOI: [10.1088/2058-9565/aa8072](https://doi.org/10.1088/2058-9565/aa8072).
- [13] Verdon, G., Broughton, M. and Biamonte, J. A quantum algorithm to train neural networks using low-depth circuits. *arXiv preprint arXiv:1712.05304*, 2017. DOI: [10.48550/arXiv.1712.05304](https://doi.org/10.48550/arXiv.1712.05304).
- [14] Schuld, M. and Killoran, N. Quantum machine learning in feature hilbert spaces. *Physical review letters*, 122(4):040504, 2019. DOI: [10.1103/physrevlett.122.040504](https://doi.org/10.1103/physrevlett.122.040504).
- [15] Benedetti, M., Garcia-Pintos, D., Perdomo, O., Leyton-Ortega, V., Nam, Y. and Perdomo-Ortiz, A. A generative modeling approach for benchmarking and training shallow quantum circuits. *npj Quantum Information*, 5(1):1–9, 2019. DOI: [10.1038/s41534-019-0157-8](https://doi.org/10.1038/s41534-019-0157-8).
- [16] Havlíček, V. et al. Supervised learning with quantum-enhanced feature spaces. *Nature*, 567(7747):209–212, 2019. DOI: [10.1038/s41586-019-0980-2](https://doi.org/10.1038/s41586-019-0980-2).
- [17] Schuld, M., Bocharov, A., Svore, K.M. and Wiebe, N. Circuit-centric quantum classifiers. *Physical Review A*, 101(3):032308, 2020. DOI: [10.1103/physreva.101.032308](https://doi.org/10.1103/physreva.101.032308).
- [18] Romero, J. and Aspuru-Guzik, A. Variational quantum generators: Generative adversarial quantum machine learning for continuous distributions. *Advanced Quantum Technologies*, 4(1):2000003, 2021. DOI: [10.1002/qute.202000003](https://doi.org/10.1002/qute.202000003).
- [19] Wecker, D., Hastings, M.B. and Troyer, M. Progress towards practical quantum variational algorithms. *Physical Review A*, 92(4):042303, 2015. DOI: [10.1103/physreva.92.042303](https://doi.org/10.1103/physreva.92.042303).
- [20] Jena, A., Genin, S. and Mosca, M. Pauli partitioning with respect to gate sets. *arXiv preprint arXiv:1907.07859*, 2019. DOI: [10.48550/arXiv.1907.07859](https://doi.org/10.48550/arXiv.1907.07859).
- [21] Verteletskyi, V., Yen, T.C. and Izmaylov, A.F. Measurement optimization in the variational quantum eigensolver using a minimum clique cover. *The Journal of chemical physics*, 152(12):124114, 2020. DOI: [10.1063/1.5141458](https://doi.org/10.1063/1.5141458).
- [22] Izmaylov, A.F., Yen, T.C., Lang, R.A. and Verteletskyi, V. Unitary partitioning approach to the measurement problem in the variational quantum eigensolver method. *Journal of chemical theory and computation*, 16(1):190–195, 2020. DOI: [10.1021/acs.jctc.9b00791](https://doi.org/10.1021/acs.jctc.9b00791).
- [23] Zhao, A., Tranter, A., Kirby, W.M., Ung, S.F., Miyake, A. and Love, P.J. Measurement reduction in variational quantum algorithms. *Physical Review A*, 101(6):062322, 2020. DOI: [10.1103/PhysRevA.101.062322](https://doi.org/10.1103/PhysRevA.101.062322).
- [24] Gokhale, P. et al. $O(N^3)$ measurement cost for variational quantum eigensolver on molecular hamiltonians. *IEEE Transactions on Quantum Engineering*, 1:1–24, 2020. DOI: [10.1109/tqe.2020.3035814](https://doi.org/10.1109/tqe.2020.3035814).
- [25] Crawford, O., van Straaten, B., Wang, D., Parks, T., Campbell, E. and Brierley, S. Efficient quantum measurement of pauli operators in the presence of finite sampling error. *Quantum*, 5:385, 2021. DOI: [10.22331/q-2021-01-20-385](https://doi.org/10.22331/q-2021-01-20-385).
- [26] Hamamura, I. and Imamichi, T. Efficient evaluation of quantum observables using entangled measurements. *npj Quantum Information*, 6(1):1–8, 2020. DOI: [10.1038/s41534-020-0284-2](https://doi.org/10.1038/s41534-020-0284-2).
- [27] McClean, J.R., Romero, J., Babbush, R. and Aspuru-Guzik, A. The theory of variational hybrid quantum-classical algorithms. *New Journal of Physics*, 18(2):023023, 2016. DOI: [10.1088/1367-2630/18/2/023023](https://doi.org/10.1088/1367-2630/18/2/023023).
- [28] Kandala, A. et al. Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets. *Nature*, 549(7671):242–246, 2017. DOI: [10.1038/nature23879](https://doi.org/10.1038/nature23879).
- [29] Jacob, F. and Ted, B. *A first course*

- in finite elements*. Wiley, 2007. DOI: [10.1002/9780470510858](https://doi.org/10.1002/9780470510858).
- [30] Kirby, W.M. and Love, P.J. Variational quantum eigensolvers for sparse hamiltonians. *Physical review letters*, 127(11):110503, 2021. DOI: [10.1103/physrevlett.127.110503](https://doi.org/10.1103/physrevlett.127.110503).
 - [31] Huang, H.Y., Kueng, R. and Preskill, J. Predicting many properties of a quantum system from very few measurements. *Nature Physics*, 16(10):1050–1057, 2020. DOI: [10.1038/s41567-020-0932-7](https://doi.org/10.1038/s41567-020-0932-7).
 - [32] Hadfield, C., Bravyi, S., Raymond, R. and Mezzacapo, A. Measurements of quantum hamiltonians with locally-biased classical shadows. *Communications in Mathematical Physics*, 2022. DOI: [10.1007/s00220-022-04343-8](https://doi.org/10.1007/s00220-022-04343-8).
 - [33] Huang, H.Y., Kueng, R. and Preskill, J. Efficient estimation of pauli observables by derandomization. *Physical Review Letters*, 127(3):030503, 2021. DOI: [10.1103/physrevlett.127.030503](https://doi.org/10.1103/physrevlett.127.030503).
 - [34] Abraham, H. et al. Qiskit: An open-source framework for quantum computing, 2019. DOI: [10.5281/zenodo.2562110](https://doi.org/10.5281/zenodo.2562110).
 - [35] de Wolf, R. Quantum computing: Lecture notes. *arXiv preprint arXiv:1907.09415*, 2019. DOI: [10.48550/arXiv.1907.09415](https://doi.org/10.48550/arXiv.1907.09415).
 - [36] McClean, J.R. et al. Openfermion: the electronic structure package for quantum computers. *Quantum Science and Technology*, 5(3):034014, 2020. DOI: [10.1088/2058-9565/ab8ebc](https://doi.org/10.1088/2058-9565/ab8ebc).
 - [37] Bron, C. and Kerbosch, J. Algorithm 457: finding all cliques of an undirected graph. *Communications of the ACM*, 16(9):575–577, 1973. DOI: [10.1145/362342.362367](https://doi.org/10.1145/362342.362367).

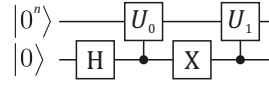


Figure S.1: Circuit for preparing $\frac{1}{\sqrt{2}}(|0\rangle|\psi_0\rangle + |1\rangle|\psi_1\rangle)$ [7], where $U_i : |0^n\rangle \mapsto |\psi_i\rangle$ for $i = 0, 1$. Note that the most significant bit is illustrated at the bottom of the circuit diagram.

A Derivation of Eq. (2)

A.1 Deriving Measurement Operators

Let $A' \in \mathbb{C}^{2^n \times 2^n}$ be an arbitrary complex matrix and $|\psi_0\rangle, |\psi_1\rangle \in \mathbb{C}^{2^n}$ be arbitrary n qubits quantum states. The expectation value $\langle\psi_0|A'|\psi_1\rangle$ can be written as

$$\langle\psi_0|A'|\psi_1\rangle = \langle\varphi|A|\varphi\rangle, \quad (\text{S.1})$$

where

$$A := \begin{cases} A' & \text{if } |\psi_0\rangle = |\psi_1\rangle \\ \begin{bmatrix} \mathbf{0} & 2A' \\ \mathbf{0} & \mathbf{0} \end{bmatrix} & \text{else} \end{cases} \quad (\text{S.2})$$

and

$$|\varphi\rangle := \begin{cases} |\psi_0\rangle = |\psi_1\rangle & \text{if } |\psi_0\rangle = |\psi_1\rangle \\ \frac{|0\rangle|\psi_0\rangle + |1\rangle|\psi_1\rangle}{\sqrt{2}} & \text{else} \end{cases}. \quad (\text{S.3})$$

Note that the state $(|0\rangle|\psi_0\rangle + |1\rangle|\psi_1\rangle)/\sqrt{2}$ can be prepared using the circuit described in Fig. S.1 [7]. A is now expressed as

$$A = \sum_{b=0}^{2^n-1} \sum_{\tilde{c}=0}^{2^n-1} A_{bc} |b\rangle\langle c| \quad (\text{S.4})$$

where

$$c := \begin{cases} \tilde{c} & \text{if } |\psi_0\rangle = |\psi_1\rangle \\ \tilde{c} + 2^n & \text{else} \end{cases}. \quad (\text{S.5})$$

Using this representation, $\langle\psi_0|A'|\psi_1\rangle$ is reformulated as

$$\langle\psi_0|A'|\psi_1\rangle = \sum_{b=0}^{2^n-1} \sum_{\tilde{c}=0}^{2^n-1} A_{bc} \langle\varphi|b\rangle\langle c|\varphi\rangle. \quad (\text{S.6})$$

Unlike existing Pauli-measurement-based methods, Pauli bases are not used to represent A in the method proposed in this study. This raises the problem that obtaining the value of $\langle\varphi|b\rangle\langle c|\varphi\rangle$ by measurement is no longer straightforward, but there is the advantage that corresponding coefficients A_{bc} can be immediately determined. Meanwhile, in the case of Pauli-measurement-based methods, the calculation of $\text{Tr}(A^\dagger \mathbf{P}_i)/2^n$ for each Pauli base $\mathbf{P}_i$ is required.

The purpose of the present method is to evaluate $\langle \varphi|b\rangle\langle c|\varphi\rangle$ for $b, c = 0, 1, \dots, 2^n - 1$ by measurement. Let $|d_{\pm}^{(b,c)}\rangle$ and $|e_{\pm}^{(b,c)}\rangle$ be defined as

$$|d_{\pm}^{(b,c)}\rangle := \frac{1}{\sqrt{2}}(|b\rangle \pm |c\rangle) \quad (\text{S.7})$$

$$|e_{\pm}^{(b,c)}\rangle := \frac{1}{\sqrt{2}}(|b\rangle \pm i|c\rangle). \quad (\text{S.8})$$

Using these definitions, the real and imaginary parts of $\langle \varphi|b\rangle\langle c|\varphi\rangle$ can be expressed as

$$\text{Re}(\langle \varphi|b\rangle\langle c|\varphi\rangle) = \frac{1}{2}(|D_+^{(b,c)}|^2 - |D_-^{(b,c)}|^2) \quad (\text{S.9})$$

$$\text{Im}(\langle \varphi|b\rangle\langle c|\varphi\rangle) = \frac{1}{2}(|E_+^{(b,c)}|^2 - |E_-^{(b,c)}|^2) \quad (\text{S.10})$$

where

$$D_{\pm}^{(b,c)} := \langle d_{\pm}^{(b,c)}|\varphi\rangle \in \mathbb{C} \quad (\text{S.11})$$

$$E_{\pm}^{(b,c)} := \langle e_{\pm}^{(b,c)}|\varphi\rangle \in \mathbb{C}. \quad (\text{S.12})$$

Equations (S.9)-(S.12) suggest that the real and imaginary parts of $\langle \varphi|b\rangle\langle c|\varphi\rangle$ can be evaluated using the probabilities of $|\varphi\rangle$ evaluated on the basis of $|d_{\pm}^{(b,c)}\rangle$ and $|e_{\pm}^{(b,c)}\rangle$, respectively. Such basis transformations can be achieved by applying the inverse of the preparation unitaries of $|d_{\pm}^{(b,c)}\rangle$ and $|e_{\pm}^{(b,c)}\rangle$ to $|\varphi\rangle$, respectively. Hence, in the following, we focus on how to obtain $|d_{\pm}^{(b,c)}\rangle$ and $|e_{\pm}^{(b,c)}\rangle$ from $|b\rangle$ and $|c\rangle$.

Let $|b\rangle$ and $|c\rangle$ be denoted by bit strings as

$$|b\rangle = |b_{n-1} \dots b_1 b_0\rangle, \quad (\text{S.13})$$

$$|c\rangle = |c_{n-1} \dots c_1 c_0\rangle, \quad (\text{S.14})$$

where $b = \sum_{i=0}^{n-1} b_i 2^i$, $c = \sum_{i=0}^{n-1} c_i 2^i$, and $b_i, c_i \in \{0, 1\}$. We define two sets of indices as

$$\mathcal{T}^{(b,c)} := \{j | b_j \neq c_j, j \in 0, 1, \dots, n-1\}, \quad (\text{S.15})$$

$$\mathcal{T}_0^{(b)} := \{j | b_j = 0, j \in 0, 1, \dots, n-1\}. \quad (\text{S.16})$$

We then denote an important index by

$$j_0^{(b,c)} \in \mathcal{T}^{(b,c)} \cap \mathcal{T}_0^{(b)}. \quad (\text{S.17})$$

Note that any element of $\mathcal{T}^{(b,c)} \cap \mathcal{T}_0^{(b)}$ can be used for $j_0^{(b,c)}$ in the following. If $\mathcal{T}^{(b,c)} \cap \mathcal{T}_0^{(b)} = \emptyset$, $\mathcal{T}^{(c,b)} \cap \mathcal{T}_0^{(c)}$ is used instead through the relationships

$$\text{Re}(\langle \varphi|b\rangle\langle c|\varphi\rangle) = \text{Re}(\langle \varphi|c\rangle\langle b|\varphi\rangle) \quad (\text{S.18})$$

$$\text{Im}(\langle \varphi|b\rangle\langle c|\varphi\rangle) = -\text{Im}(\langle \varphi|c\rangle\langle b|\varphi\rangle). \quad (\text{S.19})$$

That is, if there exists j such that $b_j \neq c_j$ but $b_j \neq 0$, b and c are exchanged using Eqs. (S.18) and (S.19). Applying the Hadamard gate to $|b\rangle$ on the $j_0^{(b,c)}$ th qubit yields

$$\begin{aligned} \text{H}(j_0^{(b,c)})|b\rangle &= \frac{1}{\sqrt{2}}|b_{n-1} \dots b_{j_0^{(b,c)}+1}\rangle(|0\rangle + |1\rangle)|b_{j_0^{(b,c)}-1} \dots b_1 b_0\rangle \\ &= \frac{1}{\sqrt{2}}(|b\rangle + |b_{n-1} \dots b_{j_0^{(b,c)}+1} c_{j_0^{(b,c)}} b_{j_0^{(b,c)}-1} \dots b_1 b_0\rangle), \end{aligned} \quad (\text{S.20})$$

where $c_{j_0^{(b,c)}} = 1$ and $H(j)$ denotes the Hadamard gate acting on the j th qubit. To convert $|b_{n-1} \cdots b_{j_0^{(b,c)}+1} c_{j_0^{(b,c)}} b_{j_0^{(b,c)}-1} \cdots b_1 b_0\rangle$ to $|c\rangle$, multiple CNOT gates are used:

$$\prod_{k \in \tilde{\mathcal{T}}^{(b,c)}} \text{CNOT}(j_0^{(b,c)}, k) H(j_0^{(b,c)}) |b\rangle = \frac{1}{\sqrt{2}} (|b\rangle + |c\rangle) = |d_+^{(b,c)}\rangle \quad (\text{S.21})$$

where

$$\tilde{\mathcal{T}}^{(b,c)} := \mathcal{T}^{(b,c)} \setminus \{j_0^{(b,c)}\} \quad (\text{S.22})$$

and $\text{CNOT}(j, k)$ denotes the CNOT gate whose control and target bits are the j th and k th qubits, respectively. When $\tilde{\mathcal{T}}^{(b,c)} = \emptyset$, the CNOT gate is not used. We now express $|d_{\pm}^{(b,c)}\rangle$ as

$$|d_{\pm}^{(b,c)}\rangle = M_{\text{Re}}^{(b,c)\dagger} |\beta_{\pm}^{(b,c)}\rangle \quad (\text{S.23})$$

where

$$|\beta_+^{(b,c)}\rangle := |b\rangle \quad (\text{S.24})$$

$$|\beta_-^{(b,c)}\rangle := X(j_0^{(b,c)}) |b\rangle = |b \oplus 2^{j_0^{(b,c)}}\rangle. \quad (\text{S.25})$$

$X(j)$ indicates the X gate acting on the j th qubit, $\oplus$ is a bit-wise exclusive OR operation, and

$$M_{\text{Re}}^{(b,c)\dagger} := \prod_{k \in \tilde{\mathcal{T}}^{(b,c)}} \text{CNOT}(j_0^{(b,c)}, k) H(j_0^{(b,c)}). \quad (\text{S.26})$$

Using the measurement operator $M_{\text{Re}}^{(b,c)}$, $|D_{\pm}^{(b,c)}|^2$ can be written as

$$|D_{\pm}^{(b,c)}|^2 = \left| \langle \beta_{\pm}^{(b,c)} | M_{\text{Re}}^{(b,c)} | \varphi \rangle \right|^2. \quad (\text{S.27})$$

This value is equivalent to the probabilities of outcomes $\beta_{\pm}^{(b,c)}$ when measuring $M_{\text{Re}}^{(b,c)} | \varphi \rangle$.

In the same sense, we define $M_{\text{Im}}^{(b,c)}$ by

$$M_{\text{Im}}^{(b,c)\dagger} := S(j_0^{(b,c)}) \prod_{k \in \tilde{\mathcal{T}}^{(b,c)}} \text{CNOT}(j_0^{(b,c)}, k) H(j_0^{(b,c)}) \quad (\text{S.28})$$

where $S(j)$ denotes the phase gate acting on the j th qubit. Then,

$$|e_{\pm}^{(b,c)}\rangle = M_{\text{Im}}^{(b,c)\dagger} |\beta_{\pm}^{(b,c)}\rangle \quad (\text{S.29})$$

and

$$|E_{\pm}^{(b,c)}|^2 = \left| \langle \beta_{\pm}^{(b,c)} | M_{\text{Im}}^{(b,c)} | \varphi \rangle \right|^2. \quad (\text{S.30})$$

A.2 Reducing the Number of Circuits

To obtain the expectation value, $\langle \psi_0 | A' | \psi_1 \rangle$, we need to measure the numerous circuits $M_{\text{Re}}^{(0,0)} | \varphi \rangle, M_{\text{Re}}^{(0,1)} | \varphi \rangle, \dots, M_{\text{Im}}^{(0,0)} | \varphi \rangle, M_{\text{Im}}^{(0,1)} | \varphi \rangle, \dots$, which requires large computational resources. Fortunately, some of these circuits can be swapped with each other. In brief, the following lemma holds.

Lemma 1 *If $b \oplus c = b' \oplus c'$, then any pair in $\{|d_+^{(b,c)}\rangle \langle d_+^{(b,c)}|, |d_-^{(b,c)}\rangle \langle d_-^{(b,c)}|, |d_+^{(b',c')}\rangle \langle d_+^{(b',c')}|, |d_-^{(b',c')}\rangle \langle d_-^{(b',c')}|\}$ is simultaneously diagonalizable. In the same way, if $b \oplus c = b' \oplus c'$, then any pair in $\{|e_+^{(b,c)}\rangle \langle e_+^{(b,c)}|, |e_-^{(b,c)}\rangle \langle e_-^{(b,c)}|, |e_+^{(b',c')}\rangle \langle e_+^{(b',c')}|, |e_-^{(b',c')}\rangle \langle e_-^{(b',c')}|\}$ is simultaneously diagonalizable.*

Proof. It is easily confirmed that $|d_{\pm}^{(b,c)}\rangle\langle d_{\pm}^{(b,c)}|$ is Hermitian for all b and c . We show that any pair in $\{|d_{+}^{(b,c)}\rangle\langle d_{+}^{(b,c)}|, |d_{-}^{(b,c)}\rangle\langle d_{-}^{(b,c)}|, |d_{+}^{(b',c')}\rangle\langle d_{+}^{(b',c')}\rangle, |d_{-}^{(b',c')}\rangle\langle d_{-}^{(b',c')}\rangle\}$ commutes. When $b \oplus c = b' \oplus c'$, the following holds.

$$\begin{aligned} b = c &\Rightarrow b' = b \oplus c \oplus c' = 0 \oplus c' = c' \\ b' = c' &\Rightarrow b = b' \oplus c' \oplus c = 0 \oplus c = c \\ b = c' &\Rightarrow b' = b \oplus c \oplus c' = 0 \oplus c = c \\ b' = c &\Rightarrow b = b' \oplus c' \oplus c = 0 \oplus c' = c' \end{aligned}$$

That is,

$$b \oplus c = b' \oplus c' \Rightarrow b = c \Leftrightarrow b' = c' \quad \text{and} \quad b = c' \Leftrightarrow b' = c.$$

From this property, we need to check if $[|b\rangle\langle c|, |b'\rangle\langle c'|] = 0$ or not, where $[A, B] = AB - BA$ is a commutator, only in the cases of (i) $b \neq b'$ and $c \neq c'$ and (ii) $b = b'$ and $c = c'$. In the same sense, we need to check if $[|b\rangle\langle c|, |c'\rangle\langle b'|] = 0$ or not only in the cases of (i) $b \neq c'$ and $c \neq b'$, (ii) $b = c'$ and $c = b'$. These checks are easily carried out and the result is that

$$b \oplus c = b' \oplus c' \Rightarrow [|b\rangle\langle c|, |b'\rangle\langle c'|] = [|b\rangle\langle c|, |c'\rangle\langle b'|] = 0.$$

As a consequence, because $|d_{\pm}^{(b,c)}\rangle = (|b\rangle \pm |c\rangle)/\sqrt{2}$,

$$b \oplus c = b' \oplus c' \Rightarrow [A, B] = 0$$

$$\text{where } A, B \in \{|d_{+}^{(b,c)}\rangle\langle d_{+}^{(b,c)}|, |d_{-}^{(b,c)}\rangle\langle d_{-}^{(b,c)}|, |d_{+}^{(b',c')}\rangle\langle d_{+}^{(b',c')}\rangle, |d_{-}^{(b',c')}\rangle\langle d_{-}^{(b',c')}\rangle\}$$

Any pair in $\{|d_{+}^{(b,c)}\rangle\langle d_{+}^{(b,c)}|, |d_{-}^{(b,c)}\rangle\langle d_{-}^{(b,c)}|, |d_{+}^{(b',c')}\rangle\langle d_{+}^{(b',c')}\rangle, |d_{-}^{(b',c')}\rangle\langle d_{-}^{(b',c')}\rangle\}$ is Hermitian and commutes, and is therefore simultaneously diagonalizable. In the same way, it can be shown that $\{|e_{+}^{(b,c)}\rangle\langle e_{+}^{(b,c)}|, |e_{-}^{(b,c)}\rangle\langle e_{-}^{(b,c)}|, |e_{+}^{(b',c')}\rangle\langle e_{+}^{(b',c')}\rangle, |e_{-}^{(b',c')}\rangle\langle e_{-}^{(b',c')}\rangle\}$ are simultaneously diagonalizable. $\square$

Lemma 1 states that there exists a unitary that simultaneously diagonalizes $|d_{\pm}^{(b,c)}\rangle\langle d_{\pm}^{(b,c)}|$ and $|d_{\pm}^{(b',c')}\rangle\langle d_{\pm}^{(b',c')}\rangle$ when $b \oplus c = b' \oplus c'$. This could appreciably reduce the number of unique circuits required to estimate the expectation value. However, lemma 1 does not state that which unitary can do. In the following, we prove that $M_{\text{Re}}^{(b,c)}$ can simultaneously diagonalize $|d_{\pm}^{(b,c)}\rangle\langle d_{\pm}^{(b,c)}|$ and $|d_{\pm}^{(b',c')}\rangle\langle d_{\pm}^{(b',c')}\rangle$ when $b \oplus c = b' \oplus c'$. In the same sense, we can show that $M_{\text{Im}}^{(b,c)}$ can simultaneously diagonalize $|e_{\pm}^{(b,c)}\rangle\langle e_{\pm}^{(b,c)}|$ and $|e_{\pm}^{(b',c')}\rangle\langle e_{\pm}^{(b',c')}\rangle$ when $b \oplus c = b' \oplus c'$.

Before proving the general case, let us consider an example. Consider the example of Eq. (S.9); i.e.,

$$\text{Re}(\langle \varphi | 011 \rangle \langle 101 | \varphi \rangle) = \frac{1}{2}(|D_{+}^{(011,101)}|^2 - |D_{-}^{(011,101)}|^2).$$

In this case, because $b = 011$ and $c = 101$,

$$\mathcal{T}^{(b,c)} \cap \mathcal{T}_0^{(b)} = \{1, 2\} \cap \{2\} = \{2\}$$

yielding

$$j_0^{(b,c)} = 2$$

as in Eq. (S.17). The bit strings used are

$$\beta_{+}^{(b,c)} = b = 011 \quad \text{and} \quad \beta_{-}^{(b,c)} = X(2)b = 111.$$

Hence, from Eq. (S.27),

$$|D_+^{(011,101)}|^2 = \left| \langle 011 | M_{\text{Re}}^{(011,101)} | \varphi \rangle \right|^2 \quad \text{and} \quad |D_-^{(011,101)}|^2 = \left| \langle 111 | M_{\text{Re}}^{(011,101)} | \varphi \rangle \right|^2,$$

where

$$M_{\text{Re}}^{(011,101)} = \text{H}(2)\text{CNOT}(2, 1).$$

We easily find that these values are equivalent to $|D_{\pm}^{(101,011)}|^2 = \left| \langle \beta_{\pm}^{(101,011)} | M_{\text{Re}}^{(101,011)} | \varphi \rangle \right|^2$, in which $M_{\text{Re}}^{(101,011)} = \text{H}(1)\text{CNOT}(1, 2)$. This means that $M_{\text{Re}}^{(011,101)}$ is no longer required to evaluate $\text{Re}(\langle \varphi | 011 \rangle \langle 101 | \varphi \rangle)$ because $M_{\text{Re}}^{(101,011)}$ can be used instead.

Here, the above example is generalized.

Theorem 1 (reprint) *If $b \oplus c = b' \oplus c'$ and $b' \neq c'$, then $M_{\text{Re}}^{(b,c)} = M_{\text{Re}}^{(b',c')}$ and $M_{\text{Im}}^{(b,c)} = M_{\text{Im}}^{(b',c')}$.*

Proof. First of all, Eq. (S.18) and (S.19) guarantee that $\langle \varphi | b \rangle \langle c | \varphi \rangle$ can be evaluated with $\langle \varphi | c \rangle \langle b | \varphi \rangle$ instead. We can therefore restrict the following discussion to the case that $b < c$. The index $j_0^{(b,c)}$ is then redefined as

$$j_0^{(b,c)} = \max_j \mathcal{T}^{(b,c)} \cap \mathcal{T}_0^{(b)}, \quad (\text{S.31})$$

which satisfies Eq. (S.17). We also consider pairs of all (b, c) and (b', c') that satisfy $b \oplus c = b' \oplus c'$. These conditions give the equality

$$\forall s \in \{\text{Re}, \text{Im}\}, \quad M_s^{(b',c')} = M_s^{(b,c)} \quad \text{for} \quad b \oplus c = b' \oplus c' \quad \text{and} \quad b' < c'. \quad (\text{S.32})$$

Equation (S.32) is proved as follows. The bit-wise exclusive OR operation, $b \oplus c$, returns bit strings that are composed of values of 1 when $b_j \neq c_j$ and 0 otherwise; e.g., $011 \oplus 101 = 110$. Therefore, the relationship $b \oplus c = b' \oplus c'$ yields $\mathcal{T}^{(b,c)} = \mathcal{T}^{(b',c')}$, given Eq. (S.15), and obviously

$$j^* = \max_j \mathcal{T}^{(b,c)} = \max_j \mathcal{T}^{(b',c')}. \quad (\text{S.33})$$

Furthermore, the conditions $b < c$ and $b' < c'$ provide $(b_{j^*}, c_{j^*}) = (b'_{j^*}, c'_{j^*}) = (0, 1)$, and thus $j^* \in \mathcal{T}_0^{(b)}$ and $j^* \in \mathcal{T}_0^{(b')}$. Consequently, we obtain $j_0^{(b,c)} = j_0^{(b',c')}$. The measurement operators $M_s^{(b,c)}$ depend only on $j_0^{(b,c)}$ and $\mathcal{T}^{(b,c)}$ (see Eqs. (S.26) and (S.28)), resulting in Eq. (S.32). In the case that $b' > c'$, as discussed above, the symmetry with respect to the exchange of b and c provides

$$\forall s \in \{\text{Re}, \text{Im}\}, \quad M_s^{(b',c')} = M_s^{(c,b)} \quad \text{for} \quad b \oplus c = b' \oplus c' \quad \text{and} \quad b' > c'. \quad (\text{S.34})$$

□

In short, by combining the properties of Eqs. (S.18) and (S.19) and Eqs. (S.32) and (S.34), all $\langle \varphi | b' \rangle \langle c' | \varphi \rangle$ that satisfy $b' \oplus c' = b \oplus c$ can be evaluated using the same measurement operator $M_s^{(b,c)}$.

A.3 Closed-form Expression of $\langle \varphi | A | \varphi \rangle$

Combining the equations derived in the previous subsections, $\langle \varphi | A | \varphi \rangle$ can be written as

$$\begin{aligned}
\langle \varphi | A | \varphi \rangle &= \sum_b A_{bb} \langle \varphi | b \rangle \langle b | \varphi \rangle \\
&\quad + \sum_b \sum_{c>b} A_{bc} \text{Re}(\langle \varphi | b \rangle \langle c | \varphi \rangle) + \sum_b \sum_{c<b} A_{bc} \text{Re}(\langle \varphi | b \rangle \langle c | \varphi \rangle) \\
&\quad + \sum_b \sum_{c>b} i A_{bc} \text{Im}(\langle \varphi | b \rangle \langle c | \varphi \rangle) + \sum_b \sum_{c<b} i A_{bc} \text{Im}(\langle \varphi | b \rangle \langle c | \varphi \rangle) \\
&= \sum_b A_{bb} \langle \varphi | b \rangle \langle b | \varphi \rangle \\
&\quad + \sum_b \sum_{c>b} \left[(A_{bc} + A_{cb}) \text{Re}(\langle \varphi | b \rangle \langle c | \varphi \rangle) + i(A_{bc} - A_{cb}) \text{Im}(\langle \varphi | b \rangle \langle c | \varphi \rangle) \right] \\
&= \sum_b A_{bb} \langle \varphi | b \rangle \langle b | \varphi \rangle + \sum_b \sum_{c>b} \left[2A_{(bc)} \text{Re}(\langle \varphi | b \rangle \langle c | \varphi \rangle) + 2iA_{[bc]} \text{Im}(\langle \varphi | b \rangle \langle c | \varphi \rangle) \right] \\
&= \sum_b A_{bb} |\langle b | \varphi \rangle|^2 \\
&\quad + \sum_b \sum_{c>b} A_{(bc)} \left(|\langle b | M_{\text{Re}}^{(b,c)} | \varphi \rangle|^2 - |\langle b \oplus 2^{j_0^{(b,c)}} | M_{\text{Re}}^{(b,c)} | \varphi \rangle|^2 \right) \\
&\quad + \sum_b \sum_{c>b} i A_{[bc]} \left(|\langle b | M_{\text{Im}}^{(b,c)} | \varphi \rangle|^2 - |\langle b \oplus 2^{j_0^{(b,c)}} | M_{\text{Im}}^{(b,c)} | \varphi \rangle|^2 \right) \tag{S.35}
\end{aligned}$$

Here, we introduce

$$l := b \oplus c. \tag{S.36}$$

Using this notation, $j_0^{(b,c)}$ can be rewritten as

$$\begin{aligned}
j_0^{(b,c)} &= \max\{j | b_j = 0, c_j = 1\} \\
&= \max\{j | b_j = 0, (l \oplus b)_j = 1\} \\
&= \begin{cases} \max\{j | l_j = 1\} & \text{for } c > b \\ \max\{j | b_j = 0, (l \oplus b)_j = 1\} & \text{for } c < b \end{cases}. \tag{S.37}
\end{aligned}$$

Note that $c = l \oplus b$. Because $c > b$ throughout Eq. (S.35), $j_0^{(b,c)}$ can be rewritten as

$$j_0^{(l)} := \max\{j | l_j = 1\}. \tag{S.38}$$

In addition, because $M_s^{(b',c')} = M_s^{(b,c)}$ when $b' \oplus c' = b \oplus c = l$, where $s \in \{\text{Re}, \text{Im}\}$, $M_s^{(b,c)}$ can be rewritten as

$$M_s^{(l)} := M_s^{(b,c)}. \tag{S.39}$$

Moreover, the following lemma holds.

Lemma 2 Let $j_0^{(l)} := \max\{j | l_j = 1\}$ and $\bar{b} := b \oplus 2^{j_0^{(l)}}$. Then, $\bar{b} \oplus l < \bar{b}$ if $b \oplus l > b$.

Proof. If $b_{j_0^{(l)}} = 1$, $b \oplus l < b$ because the operation $b \oplus l$ changes $b_{j_0^{(l)}}$ from 1 to 0, where $j_0^{(l)}$ is the most significant bit that changes value via the operation $b \oplus l$. Therefore, $b_{j_0^{(l)}} = 0$. This results in

$(b \oplus l)_{j_0^{(l)}} = 1$, $(b \oplus l \oplus 2^{j_0^{(l)}})_{j_0^{(l)}} = 0$, and $(b \oplus 2^{j_0^{(l)}})_{j_0^{(l)}} = 1$. In addition, $(b \oplus l)_{>j_0^{(l)}} = b_{>j_0^{(l)}}$ because $l_{>j_0^{(l)}} = 0$ and $x \oplus 0 = x$. This means that the magnitude relationships between $b \oplus l$ and b is determined by the value of the $j_0^{(l)}$ th bit, which is the same for the relationship between $b \oplus l \oplus 2^{j_0^{(l)}}$ and $b \oplus 2^{j_0^{(l)}}$. Therefore, $b \oplus l \oplus 2^{j_0^{(l)}} < b \oplus 2^{j_0^{(l)}}$. $\square$

According to Lemma 2,

$$\{l | b \oplus l > b\} = \{l | \bar{b} \oplus l < \bar{b}\} \quad \text{where} \quad \bar{b} := b \oplus 2^{j_0^{(l)}}, \quad (\text{S.40})$$

and obviously,

$$\{b\} = \{0, 1, \dots, 2^n - 1\} = \{\bar{b}\}. \quad (\text{S.41})$$

Using the above notations and relationships, we have

$$\begin{aligned} \sum_b \sum_{c>b} A_{(bc)} |\langle b \oplus 2^{j_0^{(b,c)}} | M_{\text{Re}}^{(b,c)} | \varphi \rangle|^2 &= \sum_b \sum_{l: b \oplus l > b} A_{(b \oplus l)} |\langle b \oplus 2^{j_0^{(l)}} | M_{\text{Re}}^{(l)} | \varphi \rangle|^2 \\ &= \sum_b \sum_{l: b \oplus l < b} A_{(b \oplus 2^{j_0^{(l)}} \oplus b \oplus l \oplus 2^{j_0^{(l)}})} |\langle b | M_{\text{Re}}^{(l)} | \varphi \rangle|^2 \end{aligned} \quad (\text{S.42})$$

and

$$\begin{aligned} \sum_b \sum_{c>b} A_{[bc]} |\langle b \oplus 2^{j_0^{(b,c)}} | M_{\text{Im}}^{(b,c)} | \varphi \rangle|^2 &= \sum_b \sum_{l: b \oplus l > b} A_{[b \oplus l]} |\langle b \oplus 2^{j_0^{(l)}} | M_{\text{Im}}^{(l)} | \varphi \rangle|^2 \\ &= \sum_b \sum_{l: b \oplus l < b} A_{[b \oplus 2^{j_0^{(l)}} \oplus b \oplus l \oplus 2^{j_0^{(l)}}]} |\langle b | M_{\text{Im}}^{(l)} | \varphi \rangle|^2. \end{aligned} \quad (\text{S.43})$$

Note that all $\bar{b}$ and b are exchanged. Substituting Eqs. (S.42) and (S.43) into (S.35), we obtain

$$\begin{aligned} \langle \varphi | A | \varphi \rangle &= \sum_b A_{bb} \langle \varphi | b \rangle \langle b | \varphi \rangle \\ &\quad + \sum_b \sum_{l \in \mathcal{S}(A) \setminus \{0^n\}} \left((-1)^w A_{(b'c')} |\langle b | M_{\text{Re}}^{(l)} | \varphi \rangle|^2 + i(-1)^w A_{[b'c']} |\langle b | M_{\text{Im}}^{(l)} | \varphi \rangle|^2 \right), \end{aligned} \quad (\text{S.44})$$

where

$$\begin{aligned} \mathcal{S}(A) &:= \{l | b \oplus l > b\} \cup \{l | b \oplus l < b\} \cup \{0^n\} \\ &= \{b \oplus c\} \end{aligned} \quad (\text{S.45})$$

and

$$(b', c', w) = \begin{cases} (b, b \oplus l, 0) & \text{for } b < b \oplus l \\ (b \oplus 2^{j_0^{(l)}}, b \oplus l \oplus 2^{j_0^{(l)}}, 1) & \text{for } b > b \oplus l \end{cases}. \quad (\text{S.46})$$

A.4 Operator Selecting Model

At first glance, the right-hand side of Eq. (S.44) does not seem to express the expectation value because it does not take the form $\sum_i a_i P(\hat{a} = a_i)$. Here, we further reformulate Eq. (S.44). During the reformulation, we introduce an operator selecting model that is useful in evaluating the variance of Eq. (S.44).

Consider that the measurement operator $M_s^{(l)}$ is selected with probability $p(l, s)$ where $l \in \mathcal{S}(A)$, $s \in \{\text{Re}, \text{Im}\}$, $M_s^{(0^n)} = I$ and $p(l = 0^n) = p(l = 0^n, s)$. The values of

$$p(l, s, b) := \begin{cases} p(l, s) |\langle b | \varphi \rangle|^2 & l = 0^n \\ p(l, s) |\langle b | M_{\text{Re}}^{(l)} | \varphi \rangle|^2 & l \neq 0^n, s = \text{Re} \\ p(l, s) |\langle b | M_{\text{Im}}^{(l)} | \varphi \rangle|^2 & l \neq 0^n, s = \text{Im} \end{cases} \quad (\text{S.47})$$

are probabilities because $p(l, s, b) \in [0, 1]$ and $\sum_{l,s,b} p(l, s, b) = 1$. Note that in this representation,

$$p(b|l, s) := \begin{cases} |\langle b | \varphi \rangle|^2 & l = 0^n \\ |\langle b | M_{\text{Re}}^{(l)} | \varphi \rangle|^2 & l \neq 0^n, s = \text{Re} \\ |\langle b | M_{\text{Im}}^{(l)} | \varphi \rangle|^2 & l \neq 0^n, s = \text{Im} \end{cases} \quad (\text{S.48})$$

Meanwhile, the coefficients

$$a(l, s, b) := \begin{cases} \frac{A_{bb}}{p(l, s)} & l = 0^n \\ \frac{(-1)^w A_{(b'c')}}{p(l, s)} & l \neq 0^n, s = \text{Re} \\ \frac{i(-1)^w A_{[b'c']}}{p(l, s)} & l \neq 0^n, s = \text{Im} \end{cases} \quad (\text{S.49})$$

can be seen as outcomes. With these definitions, Eq. (2) can be rewritten as

$$\langle \varphi | A | \varphi \rangle = \sum_{l,s,b} a(l, s, b) p(l, s, b), \quad (\text{S.50})$$

which is mathematically in the form of an expectation value.

B Derivation of Variance

According to Eq. (2), the upper bound of the variance associated with the XBM can be written as

$$\begin{aligned} \text{Var}[\hat{a}] &\leq \sum_{b=0}^{2^n-1} \left| \frac{A_{bb}}{p(l=0^n, b)} \right|^2 p(l=0^n, b) \\ &+ \sum_{l \in \mathcal{S}(A) \setminus \{0^n\}} \sum_{b=0}^{2^n-1} \left(\left| \frac{(-1)^w A_{(b'c')}}{p(l, s = \text{Re})} \right|^2 p(l, s = \text{Re}, b) + \left| \frac{i(-1)^w A_{[b'c']}}{p(l, s = \text{Im})} \right|^2 p(l, s = \text{Im}, b) \right). \end{aligned} \quad (\text{S.51})$$

When we assume

$$p(l, s) = p_{\text{uniform}}(l, s) = \frac{1}{m} \quad (\text{S.52})$$

where

$$m := |\mathcal{S}(A) \times \{\text{Re}, \text{Im}\}| \quad (\text{S.53})$$

is the number of distinct measurement operators appearing in Eq. (2), we obtain

$$\begin{aligned}
\text{Var}_{p_{\text{uniform}}}[\hat{a}] &\leq m^2 \left(\sum_{b=0}^{2^n-1} |A_{bb}|^2 p(l=0^n, b) \right. \\
&\quad \left. + \sum_{l \in \mathcal{S}(A) \setminus \{0^n\}} \sum_{b=0}^{2^n-1} \left(|A_{(b'c')}|^2 p(l, s = \text{Re}, b) + |A_{[b'c']}|^2 p(l, s = \text{Im}, b) \right) \right) \\
&\leq m^2 \max_{b,c} \{|A_{bb}|^2, |A_{(bc)}|^2, |A_{[bc]}|^2\} \\
&= m^2 \max_{b,c} \{|A_{(bc)}|^2, |A_{[bc]}|^2\}.
\end{aligned} \tag{S.54}$$

Meanwhile, when we assume

$$p(l, s) = p_{\text{weighted}}(l, s) := \begin{cases} \frac{\max_b |A_{bb}|}{Z} & l = 0^n \\ \frac{\max_b |A_{(b'c')}|}{Z} & l \neq 0^n, s = \text{Re} \\ \frac{\max_b |A_{[b'c']}|}{Z} & l \neq 0^n, s = \text{Im} \end{cases}, \tag{S.55}$$

where

$$Z := \max_b |A_{bb}| + \sum_{l \in \mathcal{S}(A) \setminus \{0^n\}} \left(\max_b |A_{(b'c')}| + \max_b |A_{[b'c']}| \right), \tag{S.56}$$

we obtain

$$\begin{aligned}
\text{Var}_{p_{\text{weighted}}}[\hat{a}] &\leq |Z|^2 \left[\sum_{b=0}^{2^n-1} \left| \frac{A_{bb}}{\max_{\tilde{b}} A_{\tilde{b}\tilde{b}}} \right|^2 p(l=0^n, b) \right. \\
&\quad \left. + \sum_{l \in \mathcal{S}(A) \setminus \{0^n\}} \sum_{b=0}^{2^n-1} \left(\left| \frac{A_{(b'c')}}{\max_{\tilde{b}'} A_{(\tilde{b}'\tilde{c}')}} \right|^2 p(l, s = \text{Re}, b) + \left| \frac{A_{[b'c']}}{\max_{\tilde{b}'} A_{[\tilde{b}'\tilde{c}]}} \right|^2 p(l, s = \text{Im}, b) \right) \right] \\
&\leq |Z|^2 \max \left\{ \left(\frac{A_{bb}}{\max_{\tilde{b}} A_{\tilde{b}\tilde{b}}} \right)^2, \left(\frac{A_{(bc)}}{\max_{\tilde{b}, \tilde{c}} A_{(\tilde{b}\tilde{c})}} \right)^2, \left(\frac{A_{[bc]}}{\max_{\tilde{b}, \tilde{c}} A_{[\tilde{b}\tilde{c}]}} \right)^2 \right\} \\
&\leq |Z|^2.
\end{aligned} \tag{S.57}$$

Obviously, $\text{Var}_{p_{\text{weighted}}}[\hat{a}] \leq \text{Var}_{p_{\text{uniform}}}[\hat{a}]$.

C Variance Comparison

C.1 Pauli measurements

In [6], A is expressed as $A = \sum_{i,j} A_{ij} |i\rangle\langle j|$, where $|i\rangle\langle j|$ is expressed using the Pauli basis as

$$|i\rangle\langle j| = \sum_k \frac{s_k^{(i,j)} \sigma_k^{(i,j)}}{2^n}, \quad s_k^{(i,j)} \in \{\pm 1, \pm i\}, \quad \sigma_k^{(i,j)} \in \{I, X, Y, Z\}^{\otimes n}, \tag{S.58}$$

where I, X, Y , and Z are the Pauli matrices. For example,

$$|010\rangle\langle 110| = |0\rangle\langle 1| \otimes |1\rangle\langle 1| \otimes |0\rangle\langle 0| = \frac{X - iY}{2} \otimes \frac{I - Z}{2} \otimes \frac{I + Z}{2}$$

$$= \frac{1}{2^3}(XII + XIZ - XZI - XZZ - iYII - iYIZ + iYZI + iYZZ).$$

The expectation value of the Pauli string can be evaluated using the Pauli measurement as

$$\langle \varphi | \sigma_k^{(i,j)} | \varphi \rangle = \sum_{b=0}^{2^n-1} t_{k,b}^{(i,j)} |\langle b | U_k^{(i,j)} | \varphi \rangle|^2, \quad t_{k,b}^{(i,j)} \in \{\pm 1\}, \quad (\text{S.59})$$

where $U_k^{(i,j)}$ is an unitary for the Pauli measurement required to estimate $\langle \varphi | \sigma_k^{(i,j)} | \varphi \rangle$, which consists of the Hadamard and phase gates. Substituting with these equations, $\langle \varphi | A | \varphi \rangle$ can be written as

$$\langle \varphi | A | \varphi \rangle = \sum_{i,j} \langle \varphi | (A_{ij} |i\rangle \langle j|) | \varphi \rangle = \sum_{i,j} \sum_k \sum_{b=0}^{2^n-1} \frac{A_{ij} s_k^{(i,j)} t_{k,b}^{(i,j)} |\langle b | U_k^{(i,j)} | \varphi \rangle|^2}{2^n} \quad (\text{S.60})$$

Assuming

$$p(i, j) = \frac{|A_{ij}|}{\sum_{p,q} |A_{pq}|}, \quad p(k|i, j) = \frac{1}{2^n}, \quad p(b|i, j, k) = |\langle b | U_k^{(i,j)} | \psi \rangle|^2, \quad (\text{S.61})$$

$\langle \varphi | A | \varphi \rangle$ can be rewritten as

$$\langle \varphi | A | \varphi \rangle = \sum_{i,j} \sum_k \sum_{b=0}^{2^n-1} \frac{A_{ij} \sum_{p,q} |A_{pq}| s_k^{(i,j)} t_{k,b}^{(i,j)}}{|A_{ij}|} p(i, j) p(k|i, j) p(b|i, j, k). \quad (\text{S.62})$$

Hence, the variance is

$$\begin{aligned} \text{Var}_p[\hat{a}] &\leq \sum_{i,j} \sum_k \sum_{b=0}^{2^n-1} \left| \frac{A_{ij} \sum_{p,q} |A_{pq}| s_k^{(i,j)} t_{k,b}^{(i,j)}}{|A_{ij}|} \right|^2 p(i, j) p(k|i, j) p(b|i, j, k) \\ &= \left(\sum_{p,q} |A_{pq}| \right)^2 \end{aligned} \quad (\text{S.63})$$

Note that the method proposed in [6] is closely related to the classical shadow with the random Pauli measurements. That is, a classical shadow, $\hat{\rho}$, associated with [6] can be expressed as

$$\text{tr}(A\hat{\rho}) = s_k^{(i,j)} t_{k,b}^{(i,j)} \sum_{p,q} |A_{pq}| \frac{A_{ij}}{|A_{ij}|}, \quad (\text{S.64})$$

which is a slightly modified version of the classical shadow with the random Pauli measurements. Concretely, [6] modifies the operator selecting model $p(i, j) p(k|i, j)$ from a uniform distribution to Eq. (S.61).

It is easily confirmed that

$$\max_b |A_{bb}| + \sum_{l \in \mathcal{S}(A) \setminus \{0^n\}} \left(\max_b |A_{(b',c')}| + \max_b |A_{[b',c']}| \right) \leq \sum_{p,q} |A_{pq}| \quad (\text{S.65})$$

because the left-hand side of Eq. (S.65), which is the same as the definition of Z (Eq. (S.56)), only sums over some elements of A whereas the right hand side of Eq. (S.65) sums over all elements of A . Thus, the upper bound of the variance associated with the XBM is tighter than that of [6] and the classical shadow with the random Pauli measurements.

C.2 Classical Shadow

The classical shadow [31] is expected to evaluate the expectation value from few measurements. In [31], two methods, namely random Pauli measurements and random Clifford measurements, were proposed. With respect to the arbitrary matrix A , the random Clifford measurements are far more suitable than the random Pauli measurements because the random Clifford measurements do not require the decomposition of A into the Pauli basis. The upper bound of the variance associated with the random Clifford measurements obeys

$$\text{Var}[\hat{a}]_{\text{shadow}} \leq \sqrt{9 + \frac{6}{2^n} \text{tr}(A^2)} = O(\text{tr}(A^2)). \quad (\text{S.66})$$

In the XBM method, the upper bound of the variance can be written as

$$\text{Var}_p[\hat{a}]_{\text{XBM}} \leq m^2 |A|_{\max}^2 \quad (\text{S.67})$$

for both models of $p(l, s)$ proposed in the main body, where $|A|_{\max} = \max_{b,c} \{|A_{bc}|\}$ is a maximum norm of A . To compare Eq. (S.66) with Eq. (S.67), we assume

$$|A|_{\max}^2 = O\left(\frac{\text{tr}(A^2)}{q}\right) \quad (\text{assumption}) \quad (\text{S.68})$$

where q is the number of non-zero elements of A . This assumption is rephrased as that the square of the maximum value of the elements of A is the same order as the mean of the square of the non-zero elements of A , and is valid for such as FEM because the element stiffness of the system is the same order in each element. Under this assumption,

$$\text{Var}_p[\hat{a}]_{\text{XBM}} = O\left(\frac{m^2}{q} \text{tr}(A^2)\right) \quad (\text{S.69})$$

This means the variance when using the XBM depends on the ratio of m^2 to q under the assumption of Eq. (S.68). The range of the number of non-zero elements, q , is

$$m \leq q \leq 2^n(2k+1) - k(k+1). \quad (\text{S.70})$$

The lower bound, $q = m$, is the case that each group for simultaneous measurement has only one element (Fig. S.2(a)) whereas the upper bound, $q = 2^n(2k+1) - k(k+1)$, is the case that all A_{ij} for $|i-j| \leq k$ have a non-zero value (Fig. S.2(b)). The worst case for the variance when using the XBM method is $q = m = 2((n - \lceil \log_2 k \rceil)k + 2^{\lceil \log_2 k \rceil}) \approx 2k(1 + n - \log_2 k)$. In this case,

$$\frac{m^2}{q} = \begin{cases} O(n^{c+1}) & \text{when } k = O(n^c) \\ O(2^n) & \text{when } k = O(2^n) \end{cases}. \quad (\text{S.71})$$

Meanwhile, when $q = 2^n(2k+1) - k(k+1) = O(k2^n)$,

$$\frac{m^2}{q} = \begin{cases} O\left(\frac{n^{c+1}}{2^n}\right) & \text{when } k = O(n^c) \\ O(1) & \text{when } k = O(2^n) \end{cases}. \quad (\text{S.72})$$

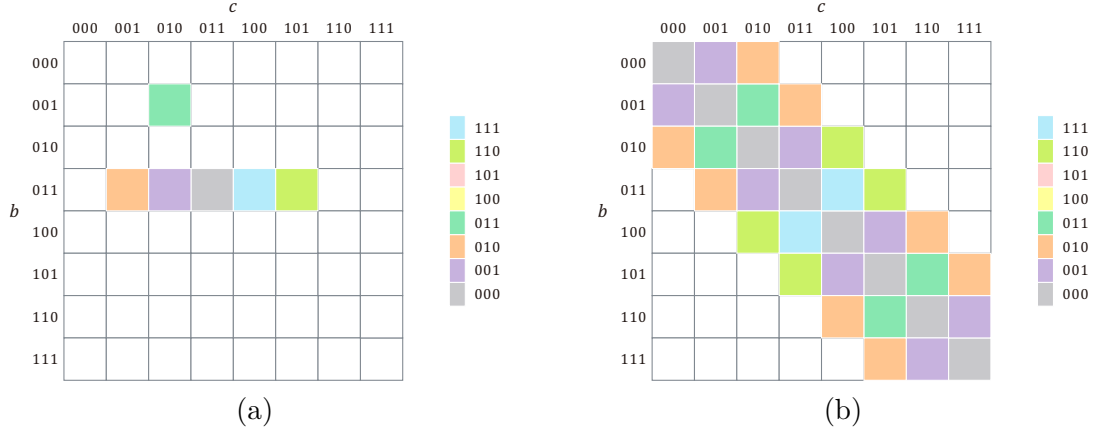


Figure S.2: Cases for the lower and upper bounds of the number of non-zero elements q for $n = 3$ qubits and $k = 2$ bandwidth. (a) Lower-bound case $q = m$. (b) Upper-bound case $q = 2^n(2k + 1) - k(k + 1)$. Components of the same color can be simultaneously evaluated by one measurement operator using the XBM method. The color range is defined by $l = b \oplus c$, where b and c are the row and column indices of A .

In summary,

$$\text{Var}_p[\hat{a}]_{\text{XBM}} = \begin{cases} O(n^{c+1} \text{tr}(A^2)) & \text{when } k = O(n^c) \text{ and } q = O(m) \\ O(2^n \text{tr}(A^2)) & \text{when } k = O(2^n) \text{ and } q = O(m) \\ O\left(\frac{n^{c+1}}{2^n} \text{tr}(A^2)\right) & \text{when } k = O(n^c) \text{ and } q = O(k2^n) \\ O(\text{tr}(A^2)) & \text{when } k = O(2^n) \text{ and } q = O(k2^n) \end{cases}. \quad (\text{S.73})$$

When $q = O(m)$, which is the case shown in Fig. S.2(a), the variance associated with the XBM is worse than that associated with the classical shadow (Eqs. (S.73)₁ and (S.73)₂). However, when $q = O(k2^n)$, which is the case shown in Fig. S.2(b), the variance associated with the XBM is at least of the same order as that associated with the classical shadow (Eqs. (S.73)₃ and (S.73)₄). In particular, when $k = O(n^c)$ and $q = O(k2^n)$, which is the case that A is a matrix with a narrow bandwidth whose non-zero elements are densely filled in the band, the variance associated with the XBM is exponentially tighter than that associated with the classical shadow (Eqs. (S.73)₃). Such A frequently appear as the global stiffness matrix in the finite element method as mentioned in the introduction of the main body.

D Classical Shadow with XBM

The XBM method is closely related to the classical shadow [31]. Let ρ , $\mathcal{U}$ and $\mathcal{M}$ be an arbitrary density matrix, unitary ensemble, and quantum channel, respectively. These are defined as

$$\mathcal{U} := \{M_s^{(l)}\}_{l \in \mathcal{S}(A), s \in \{\text{Re}, \text{Im}\}} \quad (\text{S.74})$$

and

$$\begin{aligned}\mathcal{M}(\rho) &:= \mathbb{E}_{U \sim P_{\mathcal{U}}, b \sim \langle b|\rho|b \rangle} [U^\dagger |b\rangle \langle b| U] \\ &= \sum_b \left[p(l=0^n) \langle b|\rho|b \rangle |b\rangle \langle b| + \sum_{l \in \mathcal{S}(A) \setminus \{0^n\}} \sum_{s \in \{\text{Re}, \text{Im}\}} p(l, s) \langle b|M_s^{(l)} \rho M_s^{(l)\dagger} |b\rangle M_s^{(l)\dagger} |b\rangle \langle b|M_s^{(l)} \right],\end{aligned}\tag{S.75}$$

respectively, where $P_{\mathcal{U}}$ is a probability distribution over $\mathcal{U}$. A classical shadow, $\hat{\rho} := \mathcal{M}^{-1}(U^\dagger |b\rangle \langle b| U)$, $U \sim P_{\mathcal{U}}$, $b \sim \langle b|\rho|b \rangle$, associated with the XBM method is expressed as

$$\hat{\rho}(l, s, b) = \begin{cases} \frac{|b\rangle \langle b|}{p(l=0^n)} & \text{for } l=0^n \\ (-1)^w \frac{|c'\rangle \langle b'| + |b'\rangle \langle c'|}{2p(l, s=\text{Re})} & \text{for } l \neq 0^n \text{ and } s=\text{Re} \\ i(-1)^w \frac{|c'\rangle \langle b'| - |b'\rangle \langle c'|}{2p(l, s=\text{Im})} & \text{for } l \neq 0^n \text{ and } s=\text{Im} \end{cases}\tag{S.76}$$

where (b', c', w) are defined in Table 1. The above can be easily derived by replacing A_{bc} with $|c\rangle \langle b|$ because $|\psi\rangle \langle \psi| = \sum_{b,c} |c\rangle \langle b| \langle \psi|b\rangle \langle c|\psi\rangle$ takes the same form as Eq. (1). The simple choice for the operator selecting model is

$$p(l=0^n) = p(l \neq 0^n, s) = \frac{1}{2^{n+1} - 1}\tag{S.77}$$

where $2^{n+1} - 1$ is the number of unique circuits used to evaluate the expectation value of the dense $2^n \times 2^n$ matrix. The classical shadow associated with the XBM method satisfies

$$\mathbb{E}_{l,s,b \sim p(l,s,b)} [\hat{\rho}(l, s, b)] = \rho\tag{S.78}$$

and

$$\text{tr}(A \hat{\rho}(l, s, b)) = a(l, s, b)\tag{S.79}$$

where $a(l, s, b)$ is defined in Eq. (S.49). There is a low requirement of the classical memory to store the classical shadow associated with the XBM method. That is, only the indices $\{(l, s)\}$ and bit strings measured $\{b\}$ need to be stored.

The measurement operators associated with the XBM method are Clifford circuits, which consist of CNOT, Hadamard and phase gates, and therefore the XBM can be seen as the classical shadow with the random Clifford measurement whose unitary ensemble is a subset of Clifford group $\text{Cl}(2^n)$ (Eq. (S.74)), and the unitary is selected with probability $p(l, s)$ that can be modeled to reduce the variance. From this point of view, our conclusion can be restated that by using an unitary ensemble Eq. (S.74) and appropriate operator selecting model, the expectation value can be evaluated computationally more efficiently than by using random Clifford measurements in a specific condition. Concretely,

1. the number of two-qubit entangling gates in each measurement operator is reduced from $O(n^2/\log(n))$ to $O(n)$ and
2. the variance reduces from $O(\text{tr}(A^2))$ to $O((n^{c+1}/2^n)\text{tr}(A^2))$ when non-zero elements of A are densely filled in the bandwidth k .

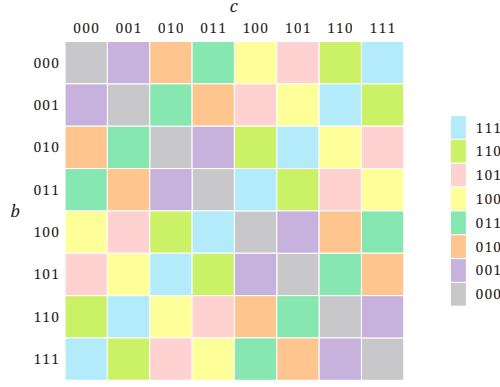


Figure S.3: Schematic illustration of the colored matrix in the case of the three-qubit system. The (b, c) component is colored with the $b \oplus c$ th color. Components of the same color can be simultaneously measured using the XBM method.

E Proofs

E.1 Colored Matrix

In proving theorems 2 and 3, we introduce a colored matrix (see Fig. S.3) as a visual reference. In the colored matrix, the (b, c) component of A is positioned on the b th row and c th column, and each component is colored with the $b \oplus c$ th color. Components of the same color are simultaneously measurable using the XBM method. Hence, the number of colors with a non-zero element value is the same as the number of unique circuits used in evaluating $\langle \psi_0 | A | \psi_1 \rangle$.

E.2 Proof of Theorem 2

We show that there exists an algorithm that produces a 1-sparse matrix including all colors in the colored matrix defined in Sec. E.1. First, for the two-qubit system, a 1-sparse matrix including all four colors is found manually (Fig. S.4(a) and (b)). Next, a $2^2 \times 2^2$ matrix is split in half and the halves are moved to the top-left and right-bottom of a $2^3 \times 2^3$ matrix as in Fig. S.4(c). The colored patterns are then copied to the top-right and bottom-left as in Fig. S.4(d). Repeating this procedure as in Fig. S.4(e) to (i), a 1-sparse matrix including all colors is obtained for any size $2^n \times 2^n$. $\square$

E.3 Proof of Theorem 3

The number of colors in a colored matrix of bandwidth k , which we denote by $N_k^{(n)}$, is counted as in Fig. S.5. From Fig. S.5, $N_k^{(n)}$ can be written as

$$N_k^{(n)} = 1 + \sum_{i=1}^k \left(a_i^{(n)} + b_i^{(n)} \right) \quad (\text{S.80})$$

where

$$a_i^{(1)} = 0, \quad b_i^{(n)} = \begin{cases} 1 & 1 \leq i \leq 2^{n-1} \\ 0 & \text{else} \end{cases}, \quad a_i^{(n)} = a_i^{(n-1)} + b_i^{(n-1)}. \quad (\text{S.81})$$

Solving this recurrence relation, we obtain

$$N_k^{(n)} = 1 + \sum_{j=1}^n \sum_{i=1}^k b_i^{(j)} = 1 + (n - \lceil \log_2 k \rceil)k + \sum_{j=1}^{\lceil \log_2 k \rceil} 2^{j-1} = (n - r)k + 2^r, \quad (\text{S.82})$$

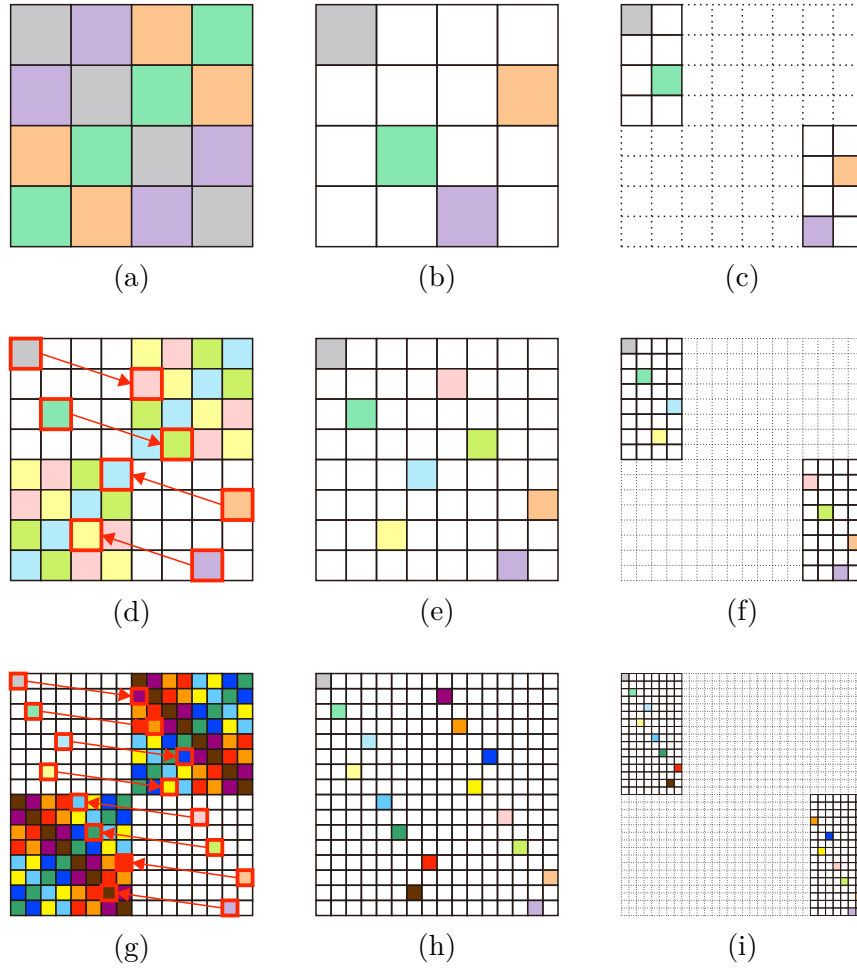


Figure S.4: Step-by-step algorithm for making a 1-sparse matrix including all colors: (a) the same illustration as in Fig. S.3 for $n = 2$; (b) in the first step, manually obtain a 1-sparse matrix for $n = 2$ including all colors; (c) split the matrix (b) in half and move the halves to the top-left and bottom-right; (d) fill the top-right and bottom-left color, copy the patterns from the top-left and bottom right to the top-right and bottom-left, respectively, and delete the unmarked square's color; (e) 1-sparse matrix for $n = 3$ including all colors; (g)-(i) repeat (c) to (f) until the matrix has the desired size.

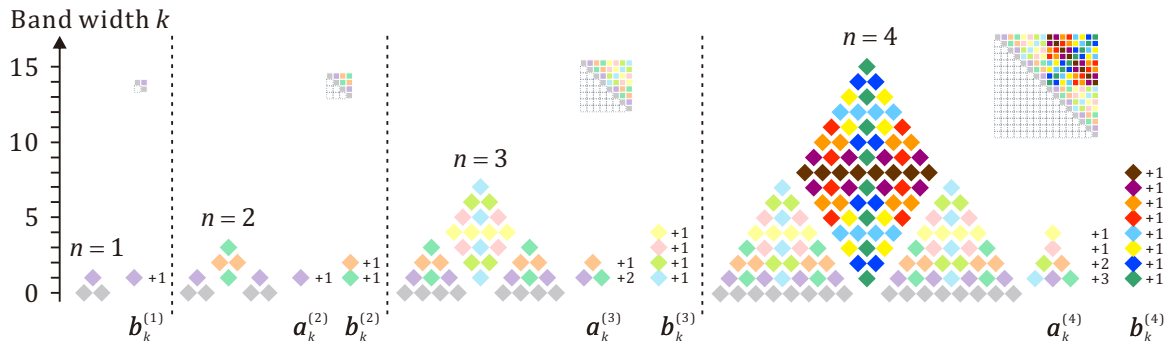


Figure S.5: Schematic illustration of the number of colors of the colored matrix for $n = 1, 2, 3, 4$.

where $r := \lceil \log_2 k \rceil$ (see Fig. S.6).

When both real and imaginary parts of $\langle \varphi | b \rangle \langle c | \varphi \rangle$ are required, the number of unique circuits is doubled. Hence, the upper bound of the number of unique circuits is $2((n-r)k + 2^r)$. When evaluating $\langle \psi_0 | A' | \psi_1 \rangle = \langle \varphi | A | \varphi \rangle$ with $|\psi_0\rangle = |\psi_1\rangle$, the matrix A can have non-zero diagonal components. Con-

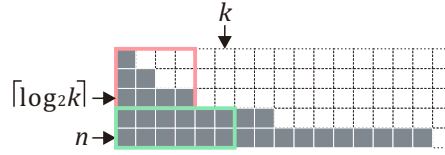


Figure S.6: Schematic illustration of matrix $b_i^{(j)}$ where the horizontal and vertical correspond to indices i and j , respectively. A filled square indicates an entry of 1 whereas a blank indicates an entry of 0. The number of entries included in the $n \times k$ square is equal to $\sum_{j=1}^n \sum_{i=1}^k b_i^{(j)}$, which can be decomposed into the red and green square corresponding to $\sum_{j=1}^{\lceil \log_2 k \rceil} 2^{j-1}$ and $(n - \lceil \log_2 k \rceil)k$, respectively.

sidering that $\text{Im}(\langle \varphi | b \rangle \langle b | \varphi \rangle) = \text{Im}(|\langle b | \varphi \rangle|^2) = 0$, the upper bound of the number of unique circuits can be reduced to $2((n - r)k + 2^r) - 1$ in this case. $\square$

E.4 Naïve Pauli measurements

When $A_{bc} \neq 0$, $\text{Tr}(A^\dagger \mathbf{P})$ takes a non-zero value where $\mathbf{P} := P_{(b \oplus c)_{n-1}} \otimes P_{(b \oplus c)_{n-2}} \otimes \cdots \otimes P_{(b \oplus c)_0}$, $P_0 \in \{I, Z\}$, $P_1 \in \{X, Y\}$, and I, X, Y and Z are the Pauli matrices. Therefore, for each distinct $b \oplus c$ such that $A_{bc} \neq 0$, 2^n Pauli strings are required to express $\langle \psi_0 | A | \psi_0 \rangle = \sum_{\mathbf{P}} (\text{Tr}(A^\dagger \mathbf{P}) / 2^n) \langle \psi_0 | \mathbf{P} | \psi_0 \rangle$. The number of distinct $b \oplus c$ is the same as the number of colors in the colored matrix defined in Sec. E.1, and the number of unique Pauli strings required to express $\langle \psi_0 | A | \psi_0 \rangle$ is therefore $2^n((n - r)k + 2^r)$. In the case that $\langle \psi_0 | A | \psi_1 \rangle$, $\mathbf{P} := P_n \otimes P_{(b \oplus c)_{n-1}} \otimes P_{(b \oplus c)_{n-2}} \otimes \cdots \otimes P_{(b \oplus c)_0}$ is used instead where $P_n \in \{X, Y\}$ and the number of unique Pauli strings required to express $\langle \psi_0 | A | \psi_1 \rangle$ is therefore $2^{n+1}((n - r)k + 2^r)$.

E.5 Evaluating Pauli strings by XBM

Consider evaluating an expectation value of one given Pauli string using the XBM method. Let $\sigma_{n-1}\sigma_{n-2}\cdots\sigma_1\sigma_0$ be a Pauli string where $\sigma_i \in \{I, X, Y, Z\}$. Converting this Pauli string into a row-column basis gives a matrix A where $A_{b,b \oplus p} \neq 0$ for all $b = 0, 1, \dots, 2^n - 1$ and p is a bit string defined as

$$p_i := \begin{cases} 0 & \text{if } \sigma_i \in \{I, Z\} \\ 1 & \text{if } \sigma_i \in \{X, Y\} \end{cases}. \quad (\text{S.83})$$

The non-zero elements of A are simultaneously measurable using the XBM method because the XBM method groups all A_{bc} that satisfy $b \oplus c = b' \oplus c'$, and $b \oplus (b \oplus p) = p$. For example, consider $XIYYZ$. In this case, $p = 10110$ and non-zero elements of A are $A_{00000,10110}, A_{00001,10111}, \dots, A_{11111,01001}$. All components of such A are simultaneously measurable using the XBM method because $00000 \oplus 10110 = 00001 \oplus 10111 = \dots = 11111 \oplus 01001 = 10110 = p$. Therefore, the number of unique circuits to evaluate $\langle \varphi | XIYYZ | \varphi \rangle$ using the XBM method is two (corresponding to the real and imaginary parts).