Classical Shadow with Decision Diagrams

Rudy Raymond[3,4,a]   Stefan Hillmich[1]   Charles Hadfield[2]   Antonio Mezzacapo[2]
Robert Wille[1,5]

Abstract: We consider the problem of estimating quantum observables on a collection of qubits, given as a linear combination of Pauli operators, with shallow quantum circuits consisting of single-qubit rotations. We introduce estimators based on classical shadow, which use decision diagrams to sample from probability distributions on measurement bases. This approach generalises previously known uniform and locally-biased classical shadows. The decision diagrams are constructed given target quantum operators and can be optimised considering different strategies. We show numerically that the estimators introduced here can produce more precise estimates on some quantum chemistry Hamiltonians, compared to previously known randomised protocols and Pauli grouping methods. The details are given at [Hillmich et al., arXiv:2105.06932]

1. Introduction

Variational quantum algorithms are based on a quantum-classical optimisation feedback loop, in which a trial parameterised quantum state is prepared on a quantum computer, a target quantum cost function is estimated on it, and a classical optimiser changes the quantum parameters to minimise the target observable. This machinery has been instrumental to find ground state energies of quantum chemistry systems, which are the smallest quantum systems believed to deliver quantum advantage in the field of quantum simulations.

Recent experiments on variational quantum algorithms (1; 2) have shown that precise estimates of complex quantum operators are essential for a successful execution of the algorithms. A finite single-qubit measurement budget can hinder the performance of the quantum-classical optimisation cycle, by introducing stochastic noise on the quantum cost function to be optimised. This problem is particularly severe for quantum chemistry systems, whose molecular Hamiltonians are composed of a linear combination of Pauli operators that grows, at worst, with the fourth power of the system size (3).

To alleviate the measurement problem, a variety of algorithms have been proposed. They all ultimately aim at obtaining precise estimations of multi-qubit quantum operators, typically given as linear combination of Pauli operators, with the smallest amount of single-qubit measurements.

The idea of using the same single-qubit measurements to estimate grouped Pauli operators that qubit-wise commute, introduced in (1) as *tensor-product basis sets*, is at the core of several measurements protocols. Some of them promise a reduction in the number of measurements for quantum chemistry systems, exploiting Pauli grouping heuristics, at the expense of an increase depth in the quantum circuits used to prepare the state to be measured (4; 5). However increased circuit depths can impair the execution on noisy quantum computers prone to decoherence. Furthermore, even in the fault-tolerant regime, bigger circuit depths can increase the overall runtime of the quantum algorithm. Addressing molecular systems, (6) shows a linear saving in the number of grouped Pauli operators when addressing molecular systems through unitary partitioning of a target Hamiltonian, while (7) finds a cubic reduction if the problem is expressed in plane wave basis, at the expense of a linear increase in circuit depth. (5) proposes *sorted insertion* to group Pauli operators based on their weights, preprocessing computations linear in the number of qubits and quadratic in the number of Pauli operators. Ref. (8) exploit simultaneous measurability of partitions of commuting Pauli strings. Exploiting the automated search for symmetries introduced (9), (10) shows a linear scaling when applied to chemistry problems, again at the expense of increased circuit depth. (11) considers random Pauli sets, and uses greedy graph coloring algorithms to determine partition of Pauli operators of Hamiltonians, conjecturing a linear saving in number of measurements if arbitrary Clifford operators before measurement are allowed.

[1]   Johannes Kepler University Linz, 4040 Linz, Austria
[2]   IBM Quantum, IBM T.J. Watson Research Center, Yorktown Heights, NY 10598
[3]   IBM Quantum, IBM Japan, 19-21 Nihonbashi Chuo-ku, Tokyo, 103-8510, Japan
[4]   Quantum Computing Center, Keio University, 3-14-1 Hiyoshi, Kohoku-ku, Yokohama, Kanagawa, 223-8522, Japan
[5]   Software Competence Center Hagenberg (SCCH) GmbH, 4232 Hagenberg, Austria
[a]   rudyhar@jp.ibm.com

Other approaches address the measurement problem while not increasing circuit depths. We refer to this specific case as the *shallow-circuit measurement problem*. It has been tackled so far formulating it in terms of graph coloring, which was solved with a variety of heuristics (10). Hybrid architectures made of quantum computers in conjunction with trained neural-network quantum states have been employed to reduce measurement variances (12).

The results contained in this work build on recent techniques based on randomised sequences of single-qubit measurements. A framework for efficiently estimating properties of reduced subsystems of quantum states was introduced in (13). There, collections of randomised measurement outcomes, labeled *classical shadows* (14), are classically stored to retrieve at a later stage expectation values of local observables. While this procedure is very well suited for retrieving many generic local observables, the uniform distribution used to draw measurement bases is not optimal in estimating with high precision specific observables such as molecular Hamiltonians. Building on this result, (15) uses *derandomisation* to deterministically change a sequence of measurement bases drawn at random, with the goal of improving precision in estimating specific sets of Pauli operators.

Improving on uniform distribution sampling, biased randomised measurement protocols can be used to improve estimate precision of given observables. The *locally-biased classical shadows* introduced in (16) are collections of random measurements generated by probability distributions optimised locally at the single-qubit level. While a bias on single-qubit product probability distributions can outperform on uniform random distributions and Pauli grouping heuristics, it still misses on improvements that can come by considering generic measurement probability distributions on a set of qubits. Here we introduce a framework to sample from probability distributions on measurement bases that generalises the local product probability distributions considered in (16). These probability distributions are generated using decision diagrams, constructed from a target quantum observable, given as a linear combination of Pauli operators.

Decision diagrams are a well-known graph-based data structure used in many disciplines of computer science to enable compact representation of data in many cases. Example applications include binary decision diagrams representing Boolean functions (17), zero-suppressed binary decision diagrams with a focus on sets (18), tagged binary decision diagrams as a combination of both (19), πDDs representing permutations (20), as well as decision diagrams representing quantum states and quantum operations (21; 22; 23). At their core, decision diagrams decompose the given data into smaller parts by successively making decisions to remove degrees of freedom, recording these decisions, and exploiting the emergence of parts that are equal. For probability distributions for the measurement problem, decision diagrams provide a natural way to bias the selection of the next measurement basis based on the previous decisions. Decision

diagrams have been used for efficient sampling on large sets by applying dynamic programming methods (24).

We propose and discuss different strategies to build decision diagrams and to optimise them in order to reduce the variance of quantum observable estimators that rely on them. We show numerically that they can improve estimation precision on some molecular Hamiltonians, compared to locally biased probability distributions. The implementation we base the results on is available `https://github.com/iic-jku/dd-quantum-measurements`.

2. The Shallow-Circuit Measurement Problem

2.1 Problem Definition

Consider an n-qubit Hamiltonian

$$H = \sum_{P \in \{I,X,Y,Z\}^n} \alpha_P P \qquad ((1))$$

with $\text{poly}(n)$ number of real coefficients α_P acting on a quantum processor, we say P is a *Pauli operator* consisting of n *single-qubit* Pauli operators and write $P = \left(\otimes_{i \in [n]} P_i\right) \in \{I,X,Y,Z\}^n$ where I, X, Y, Z are 2×2 Pauli matrices. The Hilbert space is $\mathcal{H} := (\mathbb{C}^2)^{\otimes n} = \mathbb{C}^{2^n}$. Let $\mathcal{D}(\mathcal{H})$ denote the space of quantum densities and fix some unknown $\rho \in \mathcal{D}(\mathcal{H})$. Our task is to estimate $\text{Tr}(H\rho)$ to some additive accuracy $\varepsilon > 0$.

We restrict our attention to algorithms for $\text{Tr}(H\rho)$ which are compatible with quantum processors of the current generation. Specifically we assume that the measurement bases in which we may measure ρ are of the form $B = \otimes_{i \in [n]} B_i$ where $B_i = x_i X + y_i Y + z_i Z$ and $x_i^2 + y_i^2 + z_i^2 = 1$. If we then prepare ρ many times, say $S \in \mathbb{N}$, and for each $s \in [S]$ choose a measurement basis $B^{(s)}$ in which to measure ρ we can estimate, with progressively increasing accuracy, the value of $\text{Tr}(H\rho)$. We will make two further assumptions: the choice of $B^{(s)}$ is independent of $B^{(s')}$ for $s' < s$; any such basis B is a Pauli operator $B \in \mathcal{P}^n$ where $\mathcal{P} = \{X, Y, Z\}$. The *shallow-circuit measurement problem* is how to best choose the measurement bases in order to estimate $\text{Tr}(H\rho)$ within accuracy ε with as few preparations of ρ as possible.

2.2 The General Probabilistic Measurement Framework

We can solve the shallow-circuit measurement problem by viewing it as the problem of how to best pick a probability distribution β over the measurement bases $\mathcal{P}^n$. In order to see the relationship we fix some notation. First, for a fixed Pauli operator P, let

$$\text{Cover}(P) := \{B \in \mathcal{P}^n \mid B_i = P_i \text{ whenever } P_i \neq I\}. \qquad ((2))$$

This is the set of measurement bases which allow us to estimate $\text{Tr}(P\rho)$. (We shall say that any such B *covers* P.) Next, if $B = \otimes_{i \in [n]} B_i \in \mathcal{P}^n$, then measuring qubit i in the B_i basis returns an eigenvalue $\mu(B, i) \in \{\pm 1\}$. For a subset $A \subseteq [n]$ let us declare $\mu(B, A) := \prod_{i \in A} \mu(P, i)$ with

Algorithm 1 Shallow-measurement estimation of $\mathrm{Tr}(H\rho)$ given β.

---

for shot $s \in [S]$ **do**
    Prepare ρ
    Select basis $B \in \mathcal{P}^n$ from β-distribution
    for qubit $i \in [n]$ **do**
        Measure qubit i in basis B_i giving $\mu(P, i) \in \{\pm 1\}$
    Estimate observable expectation

$$\nu^{(s)} = \sum_P \alpha_P \cdot \frac{\mathbb{1}_{\mathrm{Cover}(P)}(B)}{\zeta(P, \beta)} \cdot \mu(B, \mathrm{supp}(P))$$

  return $\nu = \frac{1}{S} \sum_{s \in [S]} \nu^{(s)}$

---

the convention that $\mu(P, \varnothing) = 1$. If we set $\mathrm{supp}(P) := \{i \in [n] | P_i \neq I\}$ then we find that $\mathrm{Tr}(P\rho)$ is estimated by $\mu(B, \mathrm{supp}(P))$ whenever B covers P. Penultimately, let $\mathbb{1}_\Omega$ represent the indicator function of a set Ω. That is $\mathbb{1}_\Omega(x)$ returns 1 if $x \in \Omega$ and 0 if $x \notin \Omega$. Finally, if $\beta : \mathcal{P}^n \to \mathbb{R}^+$ is a probability distribution then the probability that a basis B is chosen such that $\mathrm{Tr}(P\rho)$ may be estimated is

$$\zeta(P, \beta) = \sum_{B \in \mathcal{P}^n} \mathbb{1}_{\mathrm{Cover}(P)}(B) \cdot \beta(B) = \sum_{B \in \mathrm{Cover}(P)} \beta(B). \tag{3}$$

The shallow-circuit measurement problem reduces to finding $\beta : \mathcal{P}^n \to \mathbb{R}^+$ which minimises the variance of the estimator ν produced by Algorithm 1.

Let us say that β is *compatible* with the Hamiltonian H if $\zeta(P, \beta) > 0$ whenever $\alpha_P \neq 0$. In (25) we prove two results: we show that this algorithm returns an unbiased estimator for $\mathrm{Tr}(H\rho)$ provided β is compatible with H; we also calculate the variance of the estimator. The details are given at (25).

2.3 Existing Solutions and Drawbacks
2.3.1 Pauli Grouping via Graph Coloring

Pauli grouping via graph coloring has been used experimentally in (1). A detailed explanation may be found in (16, Section A.2). The core idea is to group the Pauli operators $\{P\}_{\alpha_P \neq 0}$ occurring in H into K collections and assign one measurement basis $B^{(k)} \in \mathcal{P}^n$ to each collection $k \in [K]$ such that $B^{(k)}$ allows all Pauli operators in the kth collection to be estimated. That is, if P belongs to the kth collection then $B^{(k)} \in \mathrm{Cover}(P)$.

The grouping is performed by coloring a specific graph using any graph-coloring heuristic. The graph is constructed first by assigning vertices to each Pauli operator appearing in the Hamiltonian. Second, if two vertices represent Pauli operators $P = \otimes_{i \in [n]} P_i$, and $Q = \otimes_{i \in [n]} Q_i$, then an edge is added to the graph precisely when there exists a qubit $i \in [n]$ such that $P_i, Q_i \in \mathcal{P}$ and $P_i \neq Q_i$. It follows that any vertices with the same color may be assigned a single measurement basis. Graph coloring heuristics lead to collections for which $K \ll |\mathcal{P}^n|$.

The best assignment of the probability $\beta : \{B^{(k)}\}_{k \in [K]} \to \mathbb{R}^+$ has not been rigorously studied. In (1), the measurement bases $\{B^{(k)}\}_{k \in [K]}$ are effectively sampled uniformly,

but this is due to hardware considerations. In (16, Section A.2) an improved sampling is proposed which is based on the ℓ_1 weight of the coefficients of the Pauli operators appearing in each collection. This proposal may be improved slightly by observing that some Pauli operators may be assigned to several collections. That is, sometimes, several bases from $\{B^{(k)}\}_{k \in [n]}$ may be used to estimate a single Pauli observable. This will decrease slightly the variances obtained in (16, Tables 1,2).

The main drawback of this approach is observational. That is, on Hamiltonians thus far studied in the literature, the variances are large compared to other proposals. This is despite the preprocessing steps to solve a graph coloring problem to reduce the choices of measurement bases. This may be caused by two reasons. First, the Pauli grouping stage makes no reference to the coefficients $\{\alpha_P\}$. Second, the current proposals may not be optimally assigning the distribution over $\{B^{(k)}\}_{k \in [K]}$.

2.3.2 Locally-Biased Classical Shadows

One method of choosing the distribution $\beta : \mathcal{P}^n \to \mathbb{R}^+$ has been proposed in (16) and is called locally-biased classical shadows (LBCS). It may be seen as an extension of a proposal to perform tomography, called classical shadows using random Pauli measurements (13), to the problem of estimating a single observable. LBCS fits precisely into the framework discussed in the preceding subsection. First, the class of distributions from which β may be chosen is restricted: only *product* probability distributions are considered, which we may write as $\beta = \prod_{i \in [n]} \beta_i$ where $\beta_i : \mathcal{P} \to \mathbb{R}^+$ is the probability distribution for choosing to measure the ith qubit in a basis $B_i \in \mathcal{P}$. Second, after assuming this restriction on the class of probability distributions, the choice of β is made by optimising a convex cost function associated with the Hamiltonian:

$$\mathrm{cost}_{\mathrm{diag}}(\beta) := \sum_P \alpha_P^2 \frac{1}{\prod_{i \in \mathrm{supp}(P)} \beta_i(P_i)} \tag{4}$$

Note that since β is assumed to be a product distribution, the denominator $\prod_{i \in \mathrm{supp}(P)} \beta_i(P_i)$ is precisely $\zeta(P, \beta)$. In (16), it was shown that this method leads to significant reductions in the variance of energy estimation in the context of quantum chemistry over the method of Pauli grouping via graph coloring.

Both Classical Shadows and LBCS appear to be attractive when Pauli operators in the Hamiltonian are low-weight. That is, when $|\mathrm{supp}(P)|$ is small relative to n, since in this case the denominator of Eq. (4) remains relatively small. This intuition leads to the following setup showing a shortcoming of LBCS. Consider the toy Hamiltonian $H = \otimes_{i \in [n]} X_i + \otimes_{i \in [n]} Z_i$. LBCS would assign, for each qubit, the probability $1/2$ to measure it either in the X or the Z basis. The lack of correlation in these choices implies that the only two bases X^n and Z^n which are useful for energy estimation are rarely chosen. In the parameter n it is with exponentially vanishing probability that such bases are chosen. Pauli grouping via graph coloring would per-

form much better on this example. This example motivates the search for distributions β from a wider class of probabilities. Although the preceding Hamiltonian is a toy example, there does occur Hamiltonians with similar structure. For example, for the purpose of self-testing quantum devices, the Hamiltonian $H = X_1 Z_2 \cdots Z_n + \frac{1}{n-1} \sum_{i \geq 2} Z_1 X_i$ is evaluated to test Bell-type inequalities on star graphs (26).

3. Estimators of quantum observables with Decision Diagrams

3.1 Decision Diagrams in General

Decision diagrams are a tried and tested data structure in many areas of computer science to provide a compact representation of entities in various domains. Example applications include binary decision diagrams representing conventional Boolean functions (17), zero-suppressed binary decision diagrams with a focus on sets (18), tagged binary decision diagrams as a combination of both (19), and πDDs representing permutations (20). Also in the domain of quantum computing, decision diagrams representing quantum states and quantum operations received interest (21; 22; 23) and found application, e.g., in the synthesis (27), simulation (21; 28), or verification (29) of quantum circuits.

The common idea of all decision diagram-based representations is to decompose a given original representation (e.g., of a Boolean function or quantum state) in a structured fashion that recognizes and exploits redundancies of the decomposed data in order to provide a more compact representation. The repeatedly conducted decompositions are represented by means of a directed and acyclic multi-graph, where vertices represent the decomposed data and redundancy is exploited through *shared vertices*.

Example 1. *Consider a Boolean function* $f : \{0,1\}^4 \to \{0,1\} = \bar{x}_1 \bar{x}_2 \bar{x}_3 x_4 + \bar{x}_1 x_2 x_3 \bar{x}_4 + x_1 \bar{x}_2 x_3 \bar{x}_4 + x_1 x_2 \bar{x}_3 x_4$. *A straightforward* complete *representation of this function would require the representation of a total of* 2^4 *input-output mappings, e.g., in terms of a truth table. Encoding the function as decision diagram results in a graph with only 9 nodes as illustrated in Figure 1. Here, the overall function f is first decomposed with respect to variable x_1 into two subfunctions f_6 (assuming $x_1 = 0$) and f_5 (assuming $x_1 = 1$). This is recursively continued for all remaining variables until only terminals 0 and 1 result. Whenever this decomposition yields equivalent (and, hence, redundant) sub-functions (as it is the case, e.g., for f_4), the sub-function is represented by a single shared node only—providing a more compact representation.*

3.2 Proposed Decision Diagram

In this work, we propose a type of decision diagrams aiming for a compact representation of a probability distribution over a given Hamiltonian. To this end, we first start by providing the definition of the proposed type:

Definition 1. *The decision diagram we propose is a rooted directed acyclic multi-graph $G = (V, E)$ such that all maximal directed paths consist of precisely n edges. Each edge*
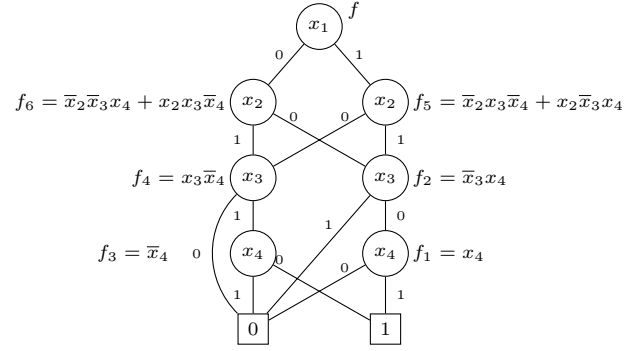

Figure 1: Decision diagram representing the Boolean function $f = \bar{x}_1 \bar{x}_2 \bar{x}_3 x_4 + \bar{x}_1 x_2 x_3 \bar{x}_4 + x_1 \bar{x}_2 x_3 \bar{x}_4 + x_1 x_2 \bar{x}_3 x_4$

$e \in E$ *is equipped with two pieces of data: A traceless Pauli operator $B(e) \in \mathcal{P}$ and a weight $w(e) \in (0,1]$ such that*

(1) *for each vertex $v \in V$, there is at most one out-going edge for each traceless Pauli operator, and*

(2) *for each vertex $v \in V$ and outgoing edges $e \in \mathrm{out}(v)$, the weights are probabilistic, i.e., $\sum_{e \in \mathrm{out}(v)} w(e) = 1$ (therefore, each vertex except for the terminal has at least one out-going edge).*

Having this structure, the edge weights in the decision diagram provide the probabilities by which a random walk should follow each edge. Multiplying the edge weight along a path gives the probability of encountering this path in a random walk. Intuitively, the probability of a path corresponds to the weight of the Pauli operators it covers. More precisely, the sum of the absolute values of the coefficients $|\alpha_P|$ with P covered by the path is used as relative probability and is encoded in the decision diagram. An example illustrates the idea.

Example 2. *Consider the Hamiltonian H for the hydrogen molecule H_2 with 4 qubits and Bravyi-Kitaev encoding, namely:*

$$
\begin{aligned}
H = &- 0.811 IIII + 0.120 IZII - 0.045 XZXI + 0.045 XIXZ \\
&+ 0.045 XIXI - 0.045 XZXZ \\
&+ 0.120 IZIZ + 0.172 ZIII - 0.225 IZZZ \\
&- 0.228 ZZII + 0.172 IIZI + 0.168 ZIZI \\
&+ 0.166 ZZZZ + 0.166 ZZZI + 0.174 ZIZZ.
\end{aligned}
$$

The full-weight terms $XZXZ$ or $ZZZZ$ cover every Pauli term in the Hamiltonian, hence, the corresponding decision diagram encoding the probability distribution only has to include these two terms.

Figure 2a illustrates a compatible decision diagram that has only two maximal paths $XZXZ$ and $ZZZZ$. The edges are labeled with a probability (edge weights of 1 are omitted for the sake of readability). Using the decision diagram, one can obtain the respectively desired probabilities by traversing the decision diagram starting at the top-most node and following the edges strictly downwards until the terminal vertex (depicted as rectangle) is reached. Generating measurements

(a) Bravyi-Kitaev   (b) Jordan-Wigner
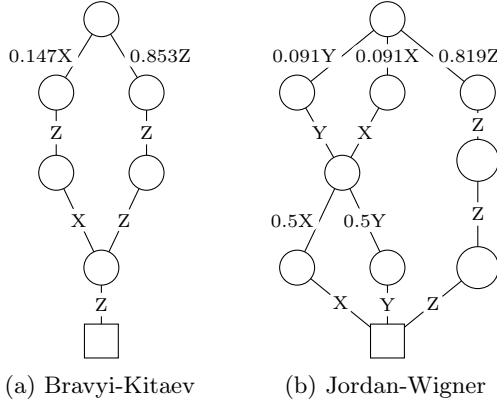
Figure 2: Decision diagrams for H_2 with 4 qubits

from this decision diagram will result in XZXZ with a probability of 0.147 or ZZZZ with 0.853.

In Figure 2a, the encoded terms end with the Z operator regardless of previous choices, hence the last decision is presented by a single vertex and edge. This may seem like a small gain, but generally larger instances have more potential for sharing.

In a similar fashion, a decision diagram representing the Jordan-Wigner encoding can be generated—yielding the structure as shown in Figure 2b.

3.3 Sampling Using the Proposed Decision Diagrams

The definition in the previous subsection does not make explicit reference to the prescribed Hamiltonian. In order to use one such instance of these decision diagrams we augment the definition with the following

Definition 2. *The decision diagram is called* compatible *with the Hamiltonian H if all Pauli observables P with $\alpha_P \neq 0$ are able to be estimated. Precisely, if $\alpha_P \neq 0$ then we require at least one directed path $(e_1, \ldots, e_n)$ such that $B \in \mathrm{Cover}(P)$ where B is the full-weight Pauli operator $\otimes_{i \in [n]} B(e_i)$.*

Let us make two observations that bring decision diagrams into the probabilistic setup of the shallow measurement problem as explained in the previous section. First, the decision diagram provides a probability distribution over full-weight Pauli operators $\beta : \mathcal{P}^n \to \mathbb{R}^+$. For $B \in \mathcal{P}^n$ such that $B = \otimes_{i \in [n]} B(e_i)$ for some maximal directed path $(e_1, \ldots, e_n)$ then we set $\beta(B) = \prod_{i \in [n]} w(e_i)$. If no such maximal directed path exists, then we set $\beta(B) = 0$. Condition 2 in Definition 1 ensures that $\sum_B \beta(B) = 1$. Indeed, using the decision diagrams proposed above, samples can be drawn by perfoming a random walk. Starting at the root vertex, a successor vertex is randomly selected according to the weights on the out-going edges. This process is repeated at the selected vertex until the terminal vertex is reached. Second, if the decision diagram is compatible with H then, for any quantum density $\rho \in \mathcal{D}(\mathcal{H})$, the estimator ν of Algorithm 1 is an unbiased estimator of the energy.

Example 2 (continued). *Consider again Figure 2a. Sampling from this decision diagram, one starts at the root ver-*

tex and randomly chooses the X or Z edge, according to the edge weights. Continuing from successor vertex of the chosen edge, the remaining decisions are fixed since each following vertex only has one out-going edge, again, resulting in either XZXZ or ZZZZ.

Decision diagrams provide a more powerful way to solve the shallow measurement problem. Indeed Pauli grouping via graph-coloring from Section 2.3 may be seen as one instance of a decision diagram according to Definition 1 and 2. Therefore any Pauli term which is covered by multiple bases under that proposal will be estimated more often under the framework here. Also, LBCS from Section 2.3 is a very simplistic instance of a decision diagram. LBCS (which observationally is better than Pauli grouping) suffers from the lack of correlation between choices of measurement bases on each qubit. The more general decision diagram framework presented here allows such correlated choices.

The larger class of distributions therefore allows us to ultimately reduce the variance associated with our estimator. Importantly, our proposal for building decision diagrams also leads to an efficient proposal for assigning weights locally such that an attractive distribution β is ultimately found. We reemphasize the possibility that such an algorithm could be extended further with the techniques of derandomisation.

3.4 Optimising the Probability Distribution on the Decision Diagram

It is a computationally difficult problem to find the optimal β. This would be the *minimum-variance unbiased estimator* (MVUE) over all such distributions $\beta : \mathcal{P}^n \to \mathbb{R}^+$. It would be interesting to understand *how close* decision diagrams get to approximating the MVUE. Nevertheless, following the method of Lagrange multipliers used in (16) to optimise the probability distribution of LBCS, we can derive similar iterative procedure to fine tune the probability distribution β based on the diagonal cost function similarly as (16). The computational cost of the iterative updates is proportional to the size of the decision diagram.

A direct implication of the use of probability-optimised decision diagrams is to generalise and improve previous results (Theorem 3 in (30) and Theorem 1 in (15)) on estimating the expectation values of a collection of Pauli operators (e.g., for partial tomography (31)) thanks to the ability to compute $\zeta(P, \beta)$ efficiently from a decision diagram. For example, the error bounds of previous results provide nontrivial bounds when the size of $\mathrm{supp}(P)$ for all Pauli P is small (or, low-weight Paulis), while those of ours can give non-trivial bounds even when some of the Paulis are of full weight. We give the details in (25).

4. Efficient Construction of the Proposed Decision Diagrams

The decision diagrams as introduced in the previous section promise to give suitable probability distribution. Still, the question remains how to efficiently transform the Hamil-

tonian consisting of coefficients and Pauli terms into a decision diagram. In the following, we describe the main ideas. Please refer to (25) for details. The full implementation is available at `https://github.com/iic-jku/dd-quantum-measurements`.

In a high-level view, the construction of the proposed decision diagram consists of multiple steps. The first stage is a preprocessing step of the Hamiltonian to reduce the number of identity-terms. The second stage is the initialisation and refinement of the decision diagram. This second stage has several steps. Only after these steps have been performed are we guaranteed that the decision diagram conforms to both Definition 1 and Definition 2. These steps include normalising the information present in the prepocessed Hamiltonian in order to maximise sharing and also removing identity-edges through *merging*. The following paragraphs explain the individual steps in more detail.

Preprocessing: Consider the Hamiltonian as presented in Eq. (1). Immediately, we remove the term $\alpha_{I^n} I^n$ as this term does not need to be estimated using the quantum processor. We also map coefficients to their absolute values ($\alpha_P \mapsto |\alpha_P|$) giving what we will call the *positive Pauli list*. Then, compatible Pauli terms in this positive Pauli list are merged to reduce the number of paths in the initial decision diagram. More precisely, for each of Pauli terms the number of compatible terms is determined. The Pauli term P_{high} with the highest number of compatible terms n_{comp} is subsequently merged into these compatible terms and the fraction $\frac{\alpha_{P_{\text{high}}}}{n_{\text{comp}}}$ is added to the coefficient of each compatible term. This merging procedure is repeated until no further merging is possible. This preprocessing provides what we shall refer to as the *reduced positive Pauli list* and shall denote it by $\mathcal{R}(H)$.

Initialisation of DD: From the reduced positive Pauli list $\mathcal{R}(H)$, an initial decision diagram whose maximal paths are all of length n is constructed. Each term in $\mathcal{R}(H)$ is associated with a unique maximal path and the coefficients of $\mathcal{R}(H)$ are assigned to the final edges in the respective path, i.e., to the edge pointing to the terminal vertex.

Normalisation of DD: Afterwards the edge weights are normalised such that the sum of weight of out-going edges equals 1. This decision diagram at this stage has sharing for common prefixes (but not suffixes) and at this point may include edges with the identity operator.

Example 3. *Consider the 4 qubit Hamiltonian for the hydrogen molecule in Jordan-Wigner encoding:*

$$
\begin{aligned}
H = &- 0.810\,IIII + 0.045\,YYXX \\
&+ 0.045\,YYYY + 0.045\,XXXX + 0.045\,XXYY \\
&+ 0.172\,ZIII - 0.225\,IZII + 0.172\,IIZI \\
&- 0.225\,IIIZ + 0.120\,ZZII + 0.168\,ZIZI \\
&+ 0.166\,ZIIZ + 0.166\,IZZI + 0.174\,IZIZ \\
&+ 0.120\,IIZZ
\end{aligned}
$$

The reduced positive Pauli list from H is

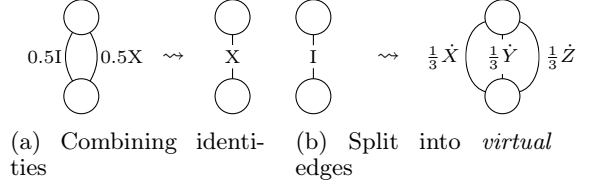(a) Combining identities (b) Split into *virtual* edges

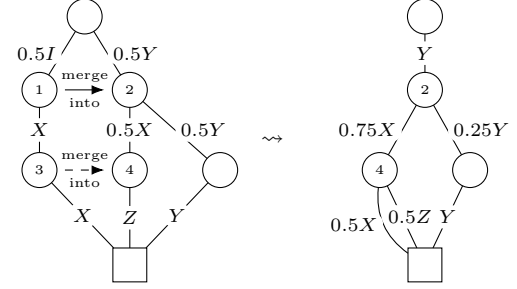Figure 3: Removing identity edges from the decision diagram

Figure 4: Merging two vertices

$$
\begin{aligned}
\mathcal{R}(H) =&\,0.045\,YYXX + 0.045\,YYYY + 0.045\,XXXX \\
&+ 0.045\,XXYY + 1.714\,ZZZZ
\end{aligned}
$$

Before considering removing potentially remaining identity edges in the decision diagram, functionally equivalent vertices are merged. Two vertices are equivalent if they have the same successors considering the Pauli operator and respective weight. Given a suitable hash function, finding equivalent nodes is linear in the number of nodes (23).

After the previous steps, the decision diagram may still have identity edges, which have to be removed to get a proper probability distribution over the Hamiltonian. The potentially remaining identity edges are removed in three steps.

The first two steps are local operations. Given two fixed vertices $u, v \in V$ the following checks are performed:

(1) For fixed $u, v \in V$, if there is an edge $u \xrightarrow{I} v$ and any $u \xrightarrow{\{X,Y,Z\}} v$: Remove $u \xrightarrow{I} v$ and add the weight to the remaining edge from $u \xrightarrow{\{X,Y,Z\}} v$ with the smallest weight. Figure 3a illustrates this step in an example with one further edge.

(2) Again, for fixed $u, v \in V$, if there is only one edge $u \xrightarrow{I} v$: Split $u \xrightarrow{I} v$ into *virtual edges* (denoted by a dot above the operator) $u \xrightarrow{\dot{X}} v$, $u \xrightarrow{\dot{Y}} v$, and $u \xrightarrow{\dot{Z}} v$ with weights $\frac{1}{3}$ each. Figure 3b illustrates this step.

The remaining identity edges cannot be removed by only considering individual pairs of nodes, but require a more global approach. Recall that at this point there are no two vertices with only an identity edge between them. So for $u \xrightarrow{I} v$ and any $u \xrightarrow{\{X,Y,Z\}} v'$ we merge v into v' and adjust the weights accordingly. More precisely, the merging is handled by checking the following list for each out-going edge of v:

(1) If the target vertex v' does not have an out-edge with

the same Pauli operator as the currently considered out-edge of v, add this edge to v'.

(2) If the currently considered out-edge of v and the out-edge of v' with the same operator point to same vertex (which may be the terminal vertex), the weights stay the same.

(3) Otherwise the merging process has to recurse to merge the successors of v and v' with corresponding Pauli operators.

During the merging of two edges, the resulting edge is only *virtual* if both previous edges were virtual. Performing the merging process from the terminal vertex upwards ensures that the recursive applications never encounter an identity edge. At this point there may be superfluous virtual edges left, which are removed to reduce the number of paths. After the merging process is completed, the decision diagram is renormalised to ensure the sum of weights of out-going edges on a node equals 1. An example illustrates the idea.

Example 4. *Consider the left-hand side of Figure 4. This decision diagram has a top vertex with two out-going edges with the operators I and Y. To remove the identity edge, the target vertex of the identity edge (1) is merged into the target vertex of the Pauli-Y edge (2) (indicated by a solid an arrow labeled "merge into"). Since both (1) and (2) have an out-going Pauli-X edge to different target vertices, theses targets (3) and (4) have to merged as well in a recursive fashion (indicated by a dashed arrow).*

Finally, the experimental evaluation on decision diagrams constructed as described above and the comparison against state-of-the-art methods can be found in (25).

5. Conclusion

We have introduced a new estimator for measuring quantum operators defined as linear combination of tensor products of single-qubit Pauli operators. The estimator is defined within a probabilistic measurement framework, where single-qubit measurement bases are drawn from probability distributions obtained using decision diagrams. The decision diagrams used to sample from measurement bases are constructed from target quantum operators, typically Hamiltonians, by associating paths in the diagrams with Pauli operators present in the Hamiltonians. The diagrams can then be simplified by removing paths with identities operators, and merging equivalent sub-paths.

We have shown that representing probability distributions with decision diagrams generalises previous classical-shadow randomises approaches to the measurement problem, namely the uniform (13) and the locally-biased one (16). This generalisation comes with additional degrees of freedom that characterize each diagram, and introduce correlations between measurement bases for each qubit. We presented different strategies to optimise these additional degrees of freedom, and have shown numerically in (25) that they can outperform locally-biased approaches as well as

Pauli grouping strategies, on selected molecular Hamiltonian models. We foresee that future refined approaches in the construction and optimisation of the diagrams could further improve on the improvements in estimation precision reported here, especially considering problem-specific decision diagram construction methods.

References

[1] Kandala, A., Mezzacapo, A., Temme, K., Takita, M., Brink, M., Chow, J. M. and Gambetta, J. M.: Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets, *Nature*, Vol. 549, No. 7671, p. 242 (2017).

[2] Hempel, C., Maier, C., Romero, J., McClean, J., Monz, T., Shen, H., Jurcevic, P., Lanyon, B. P., Love, P., Babbush, R. et al.: Quantum chemistry calculations on a trapped-ion quantum simulator, *Phys. Rev. X*, Vol. 8, No. 3, p. 031022 (2018).

[3] Wecker, D., Hastings, M. B. and Troyer, M.: Progress towards practical quantum variational algorithms, *Phys. Rev. A*, Vol. –, pp. – (online), DOI: 10.1103/PhysRevA.92.042303 (2015).

[4] 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*, Vol. 16, No. 1, pp. 190–195 (2019).

[5] Crawford, O., Straaten, B. v., Wang, D., Parks, T., Campbell, E. and Brierley, S.: Efficient quantum measurement of Pauli operators in the presence of finite sampling error, *Quantum*, Vol. 5, p. 385 (online), DOI: 10.22331/q-2021-01-20-385 (2021).

[6] Zhao, A., Tranter, A., Kirby, W. M., Ung, S. F., Miyake, A. and Love, P. J.: Measurement reduction in variational quantum algorithms, *Phys. Rev. A*, Vol. 101, No. 6, p. 062322 (2020).

[7] Huggins, W. J., McClean, J. R., Rubin, N. C., Jiang, Z., Wiebe, N., Whaley, K. B. and Babbush, R.: Efficient and noise resilient measurements for quantum chemistry on near-term quantum computers, *npj Quantum Information*, Vol. 7, No. 1, pp. – (2021).

[8] Hamamura, I. and Imamichi, T.: Efficient evaluation of quantum observables using entangled measurements, *npj Quantum Information*, Vol. 6, No. 1, pp.

1–8 (2020).

[9] Bravyi, S., Gambetta, J. M., Mezzacapo, A. and Temme, K.: Tapering off qubits to simulate fermionic Hamiltonians (2017). arXiv:1701.08213.

[10] Yen, T.-C., Verteletskyi, V. and Izmaylov, A. F.: Measuring all compatible operators in one series of single-qubit measurements using unitary transformations, *Journal of chemical theory and computation*, Vol. 16, No. 4, pp. – (2020).

[11] Jena, A., Genin, S. and Mosca, M.: Pauli Partitioning with Respect to Gate Sets (2019). arXiv:1907.07859.

[12] Torlai, G., Mazzola, G., Carleo, G. and Mezzacapo, A.: Precise measurement of quantum observables with neural-network estimators, *Phys. Rev. Research*, Vol. 2, p. 022060 (online), DOI: 10.1103/PhysRevResearch.2.022060 (2020).

[13] Huang, H.-Y., Kueng, R. and Preskill, J.: Predicting many properties of a quantum system from very few measurements, *Nature Physics*, Vol. –, pp. – (online), DOI: 10.1038/s41567-020-0932-7 (2020).

[14] Aaronson, S.: Shadow tomography of quantum states, *SIAM Journal on Computing*, Vol. 49, No. 5, pp. STOC18–368 (2020).

[15] Huang, H.-Y., Kueng, R. and Preskill, J.: Efficient estimation of Pauli observables by derandomization (2021). arXiv:2103.07510.

[16] Hadfield, C., Bravyi, S., Raymond, R. and Mezzacapo, A.: Measurements of Quantum Hamiltonians with Locally-Biased Classical Shadows (2020). arXiv:2006.15788.

[17] Bryant, R. E.: Symbolic manipulation of Boolean functions using a graphical representation, *Design Automation Conference*, pp. 688–694 (online), DOI: 10.1145/317825.317964 (1985).

[18] Minato, S.: Zero-suppressed BDDs and their applications, *Int. J. Softw. Tools Technol. Transf.*, Vol. 3, No. 2, pp. 156–170 (online), DOI: 10.1007/s100090100038 (2001).

[19] van Dijk, T., Wille, R. and Meolic, R.: Tagged BDDs: Combining reduction rules from different decision diagram types, *Formal Methods in Computer Aided Design*, pp. 108–115 (online), DOI: 10.23919/FMCAD.2017.8102248 (2017).

[20] Minato, S.: πDD: A New Decision Diagram for Efficient Problem Solving in Permutation Space, *Theory and Applications of Satisfiability Testing*, Lecture Notes in Computer Science, Vol. 6695, Springer, pp. 90–104 (online), DOI: 10.1007/978-3-642-21581-0_9 (2011).

[21] Viamontes, G. F., Markov, I. L. and Hayes, J. P.: *Quantum Circuit Simulation*, Springer (2009).

[22] Niemann, P., Wille, R., Miller, D. M., Thornton, M. A. and Drechsler, R.: QMDDs: Efficient Quantum Function Representation and Manipulation, *IEEE Trans. Comput. Aided Des. Integr. Circuits Syst.*, Vol. 35, No. 1, pp. 86–99 (2016).

[23] Zulehner, A., Hillmich, S. and Wille, R.: How to Efficiently Handle Complex Values? Implementing Decision Diagrams for Quantum Computing, *International Conference on Computer-Aided Design*, (online), DOI: 10.1109/ICCAD45719.2019.8942057 (2019).

[24] Sakaue, S., Ishihata, M. and Minato, S.: Efficient Bandit Combinatorial Optimization Algorithm with Zero-suppressed Binary Decision Diagrams, *Proceedings of the Twenty-First International Conference on Artificial Intelligence and Statistics*, Proceedings of Machine Learning Research, Vol. 84, pp. 585–594 (online), available from ⟨http://proceedings.mlr.press/v84/sakaue18a.html⟩ (2018).

[25] Hillmich, S., Hadfield, C., Raymond, R., Mezzacapo, A. and Wille, R.: Decision Diagrams for Quantum Measurements with Shallow Circuits (2021).

[26] Yang, B., Raymond, R., Imai, H., Chang, H. and Hiraishi, H.: Testing Scalable Bell Inequalities for Quantum Graph States on IBM Quantum Devices (2021). arXiv:2101.10307.

[27] Zulehner, A. and Wille, R.: One-Pass Design of Reversible Circuits: Combining Embedding and Synthesis for Reversible Logic, *IEEE Trans. Comput. Aided Des. Integr. Circuits Syst.*, Vol. 37, No. 5, pp. 996–1008 (2018).

[28] Zulehner, A. and Wille, R.: Advanced Simulation of Quantum Computations, *IEEE Trans. Comput. Aided Des. Integr. Circuits Syst.*, Vol. 38, No. 5, pp. 848–859 (2019).

[29] Burgholzer, L. and Wille, R.: Advanced Equivalence Checking for Quantum Circuits, *IEEE Trans. Comput. Aided Des. Integr. Circuits Syst.*, Vol. –, pp. – (2021).

[30] Evans, T. J., Harper, R. and Flammia, S. T.: Scalable Bayesian Hamiltonian learning (2019). arXiv:1912.07636.

[31] Cotler, J. and Wilczek, F.: Quantum overlapping tomography, *Phys. Rev. Lett.*, Vol. 124, No. 10, p. 100401 (2020).