Classical Shadows

Charles Hadfield

July 7, 2021

IBM **Quantum**

Outline

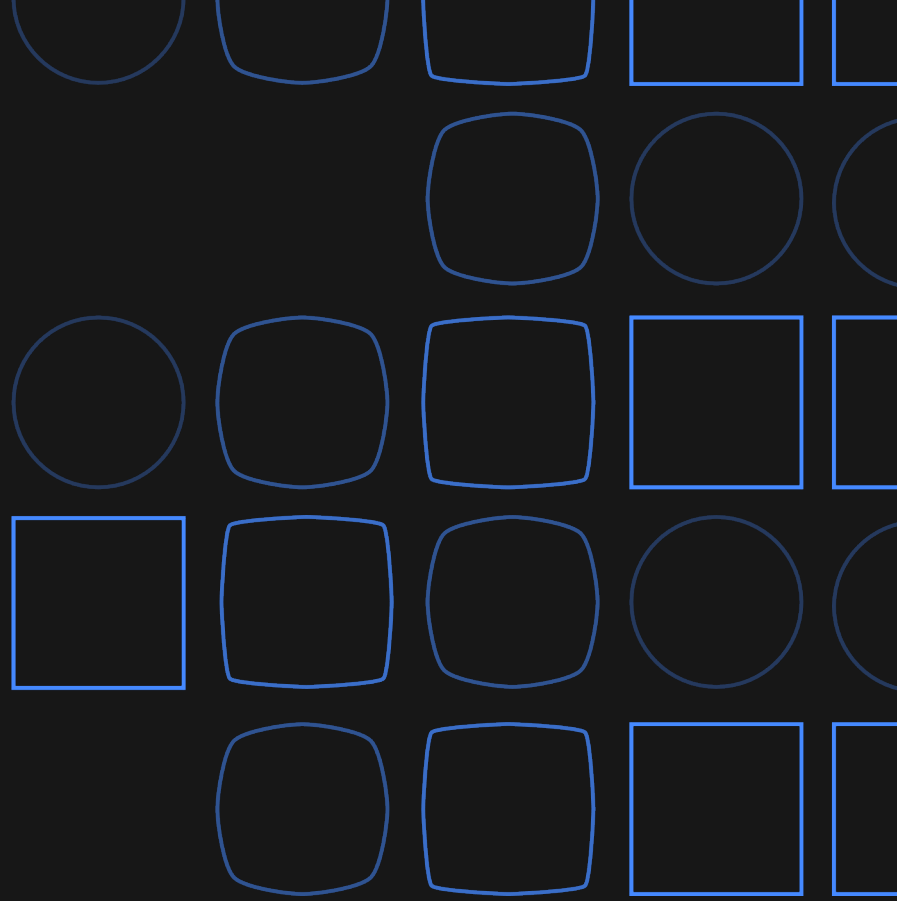
Classical shadows for tomography

- Shadow tomography
- Classical shadows

The shallow-circuit measurement problem

- Soln 0: Ell-1 sampling
- Soln 0: LDF grouping
- Soln 0: classical shadows
- Soln 1: **locally-biased classical shadows**
- Soln 3: **globally-biased classical shadows**
- Soln 2: derandomized classical shadows
- Soln 5: overlapped grouping method
- Soln 4: **adaptive Pauli shadows**

Classical shadows



nature physics

ARTICLES

https://doi.org/10.1038/s41567-020-0932-7

Check for updates

Predicting many properties of a quantum system from very few measurements

Hsin-Yuan Huang[1,2] ✉, Richard Kueng[1,2,3] and John Preskill[1,2,4]

Predicting the properties of complex, large-scale quantum systems is essential for developing quantum technologies. We present an efficient method for constructing an approximate classical description of a quantum state using very few measurements of the state. This description, called a 'classical shadow', can be used to predict many different properties; order $\log(M)$ measurements suffice to accurately predict M different functions of the state with high success probability. The number of measurements is independent of the system size and saturates information-theoretic lower bounds. Moreover, target properties to predict can be selected after the measurements are completed. We support our theoretical findings with extensive numerical experiments. We apply classical shadows to predict quantum fidelities, entanglement entropies, two-point correlation functions, expectation values of local observables and the energy variance of many-body local Hamiltonians. The numerical results highlight the advantages of classical shadows relative to previously known methods.

Classical shadows

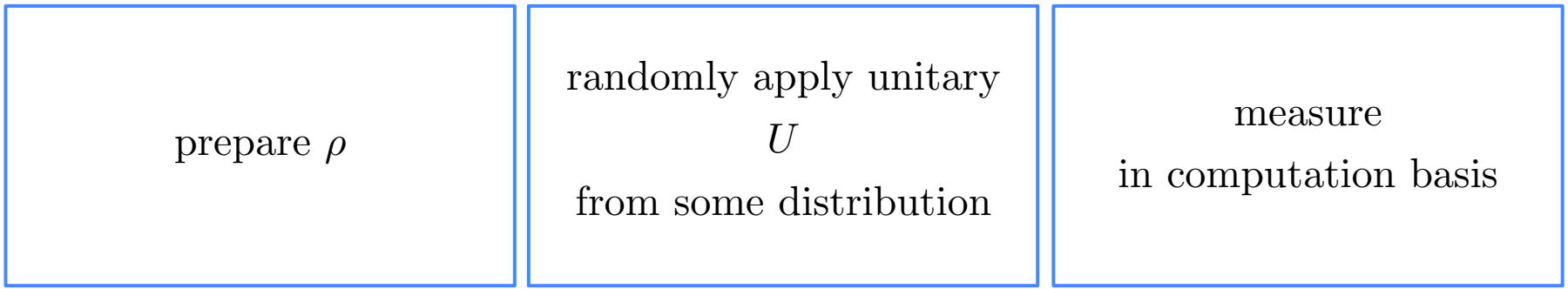
Given an *n* qubit state and *M* expectation values to estimate, how many copies of the state are required?

Today's notion of a "shadow":

Classical shadows

Given an *n* qubit state and *M* expectation values to estimate, how many copies of the state are required?

Today's notion of a "shadow":

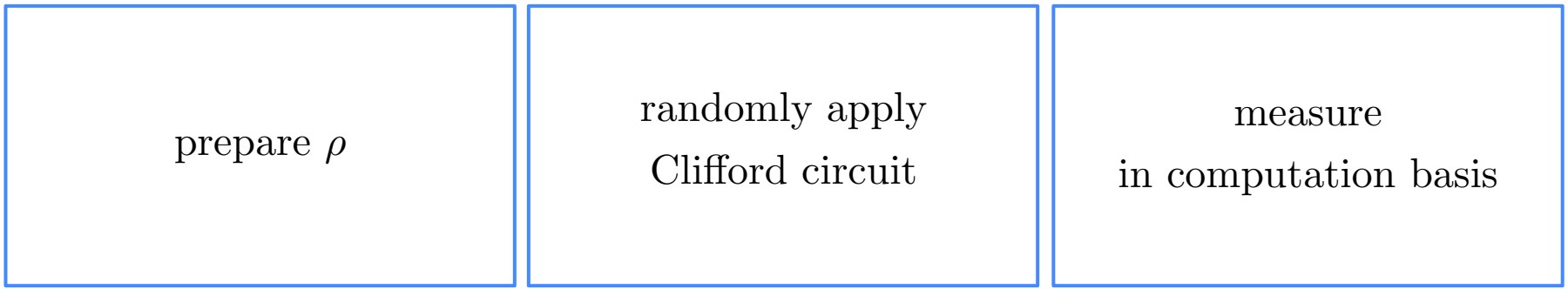


And record the outcomes in a classically-efficient data structure.

Random Clifford measurements

Given an *n* qubit state and *M* expectation values to estimate, how many copies of the state are required?

Today's notion of a "shadow":

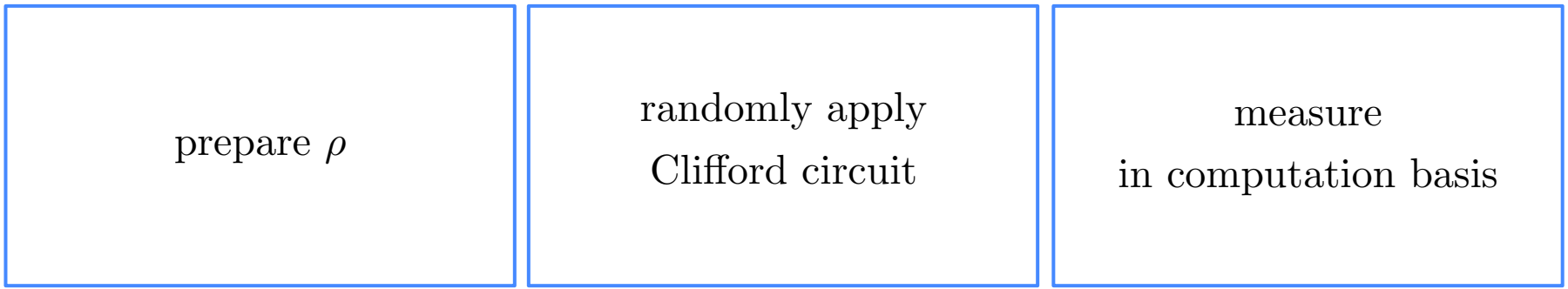


And record the outcomes in a classically-efficient data structure.

Random Clifford measurements

Given an n qubit state and M expectation values to estimate, how many copies of the state are required?

Today's notion of a "shadow":

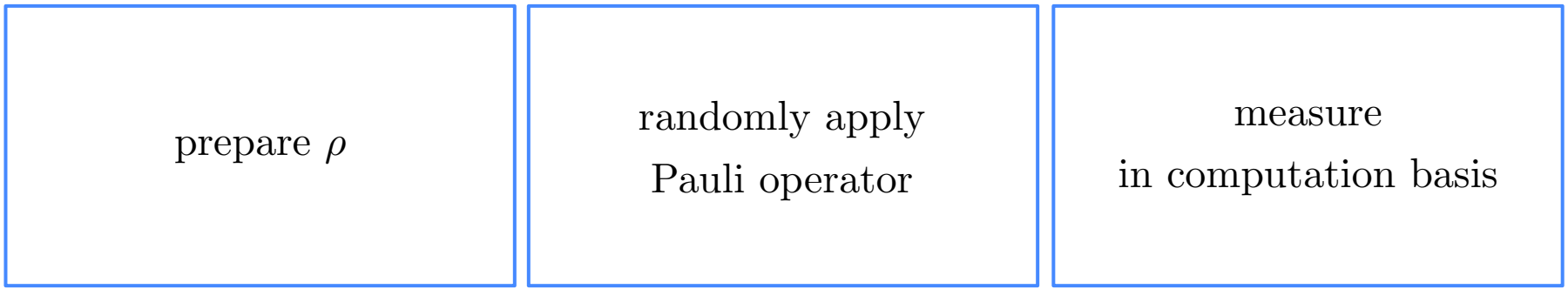


And record the outcomes in a classically-efficient data structure. (Symplectic representation over F_2)

Random Pauli measurements

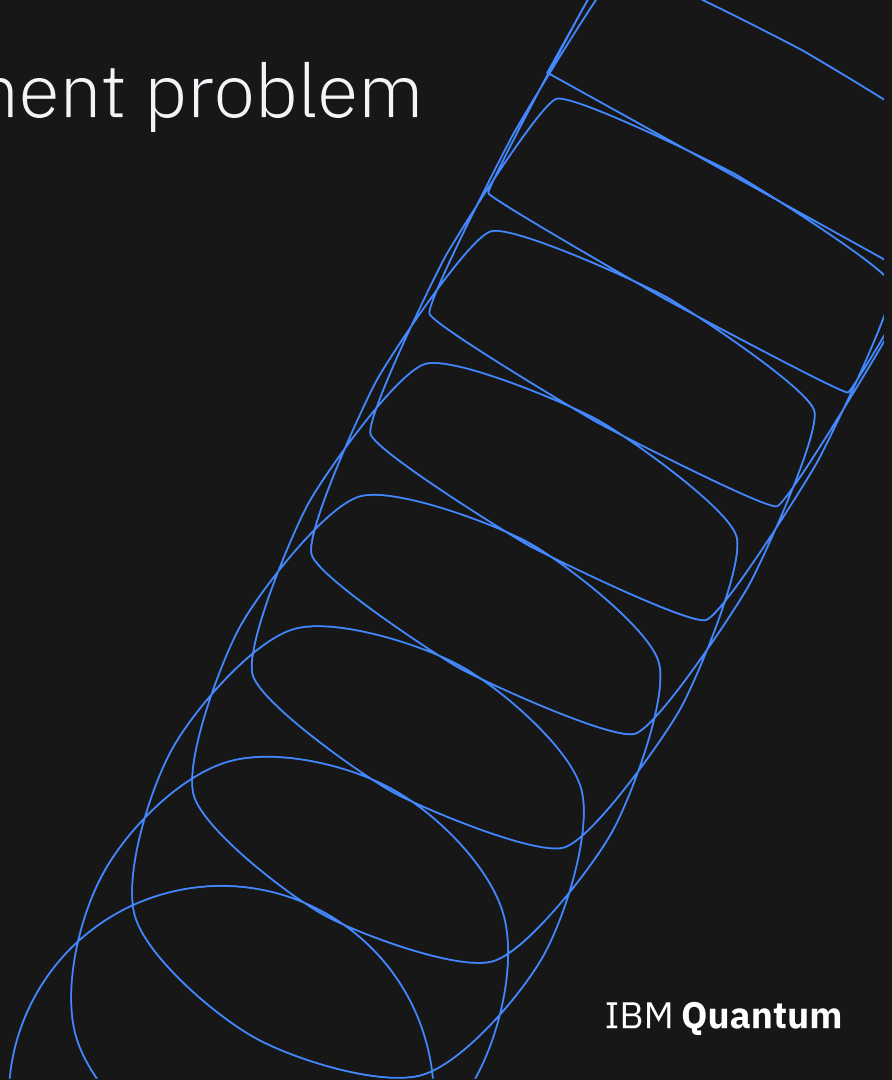
Given an n qubit state and M expectation values to estimate, how many copies of the state are required?

Today's notion of a "shadow":



And record the outcomes in a classically-efficient data structure. (Symplectic representation over F_2)

The shallow-circuit measurement problem



Shallow circuits for VQE

Our Hamiltonian will be on *n* qubits

$$H = \sum_P \alpha_P P \qquad P = \otimes_{i\in[n]} P_i \qquad P_i \in \{I, X, Y, Z\}$$

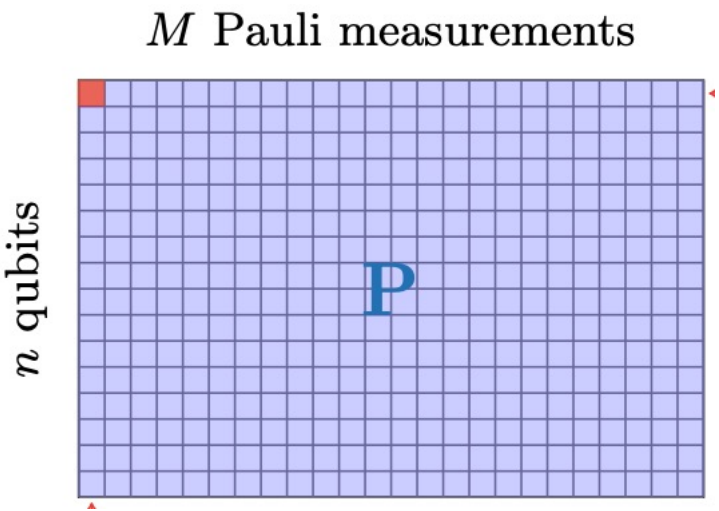
Once a state rho has been prepared, *no* entangling gates may be applied.

Measurement bases will be $B = \otimes_{i\in[n]} B_i \qquad B_i \in \{X, Y, Z\}$

How does one best choose the choice of measurement bases in order to estimate the energy, given the state, to accuracy epsilon, with *as few as possible* preparations of the state?

Algorithm 1 Estimation of Energy

for measurement $m \in [M]$ **do**
  Prepare state ρ
  Choose measurement basis B according to your algorithm
  Measure ρ in basis and estimate Pauli operators
return energy estimate



Shallow circuits for VQE

Every algorithm's goal will be to build an *unbiased* estimator

$$\mathbb{E}(\nu) = \mathrm{Tr}(H\rho)$$

whose single-shot variance is proportional to the required number of preparations of the state

$$\varepsilon = \sqrt{\frac{\mathrm{Var}(\nu)}{M}}$$

Shallow circuits for H_2O on 14 qubits

A back of the envelope demonstration of time-advantages:

Chemical accuracy requires additive-error accuracy of ~ 1 mHartree

So if I finally get a candidate density close to the true ground state, how many shots ($x10^6$) are required to get an average error of 1 mHartree?

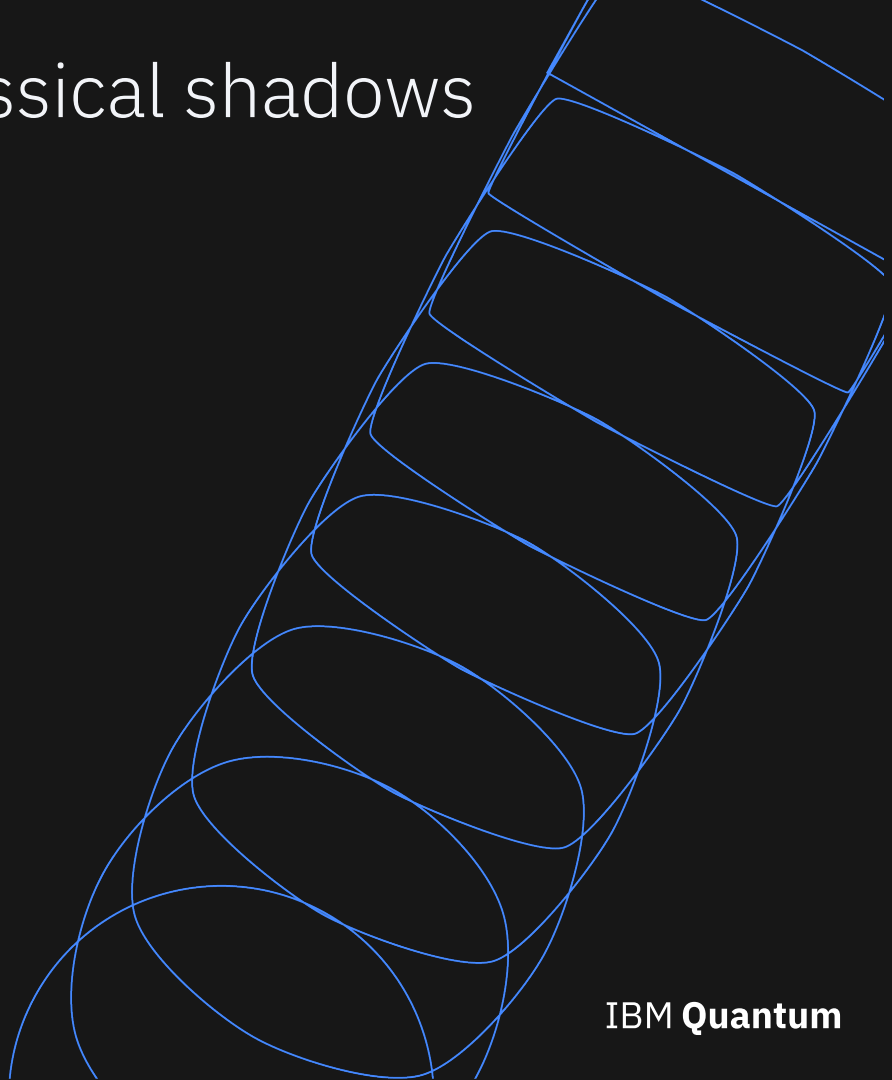
Ell-1 sampling (2015)	4 400
LDF grouping* (2017)	1 000
Classical shadows (2020)	2 800
Locally-biased CS (2020)	250
Globally-biased CS (2021)	300
Derandomized CS (2021)	15
Adaptive Pauli shadows (2021)	12
Overlapped grouping (2021)	16

Multiply this by how many different guesses in VQE are required before reaching an accurate representation of the ground state…

Other molecules

		Qiskit	LBCS	GBCS	Caltech	APS
LiH	JW	55	15	10	1	1.6
	P	85	30	15	1	2.5
	BK	75	70	15	1.6	5
BeH_2	JW	140	70	30	3.6	3.6
	P	240	130	40	8.1	3.6
	BK	200	240	60	3.6	3.6
H_2O	JW	1000	260	300	14	12
	P	2700	430	430	50	12
	BK	2100	1400	530	40	10
NH_3	JW	900	350		32	16
	P	2600	630		44	19
	BK	2150	380		14	12

Solution 1: Locally-biased classical shadows



IBM **Quantum**

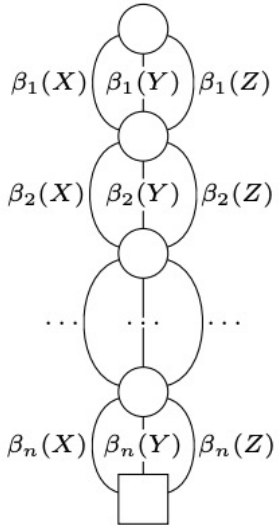
Locally-biased classical shadows

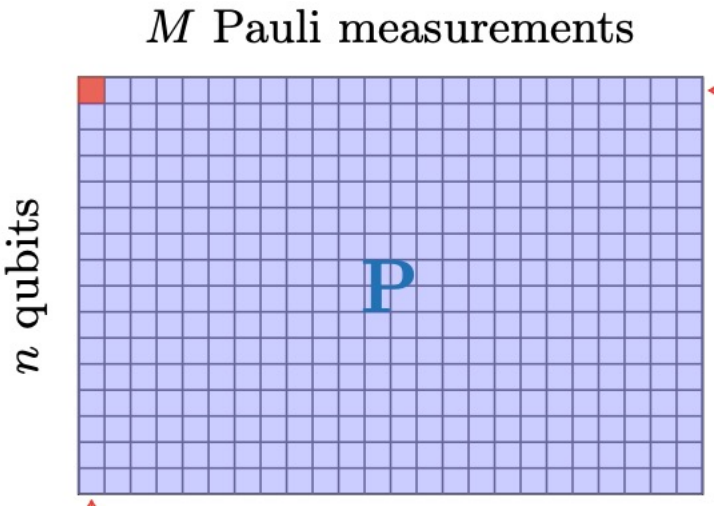
MEASUREMENTS OF QUANTUM HAMILTONIANS WITH LOCALLY-BIASED CLASSICAL SHADOWS

CHARLES HADFIELD, SERGEY BRAVYI, RUDY RAYMOND, AND ANTONIO MEZZACAPO

ABSTRACT. Obtaining precise estimates of quantum observables is a crucial step of variational quantum algorithms. We consider the problem of estimating expectation values of molecular Hamiltonians, obtained on states prepared on a quantum computer. We propose a novel estimator for this task, which is locally optimised with knowledge of the Hamiltonian and a classical approximation to the underlying quantum state. Our estimator is based on the concept of classical shadows of a quantum state, and has the important property of not adding to the circuit depth for the state preparation. We test its performance numerically for molecular Hamiltonians of increasing size, finding a sizable reduction in variance with respect to current measurement protocols that do not increase circuit depths.

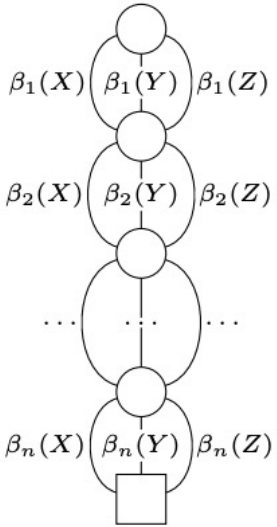
How do we choose our measurement bases?





How do we choose our measurement bases?

IBM **Quantum**



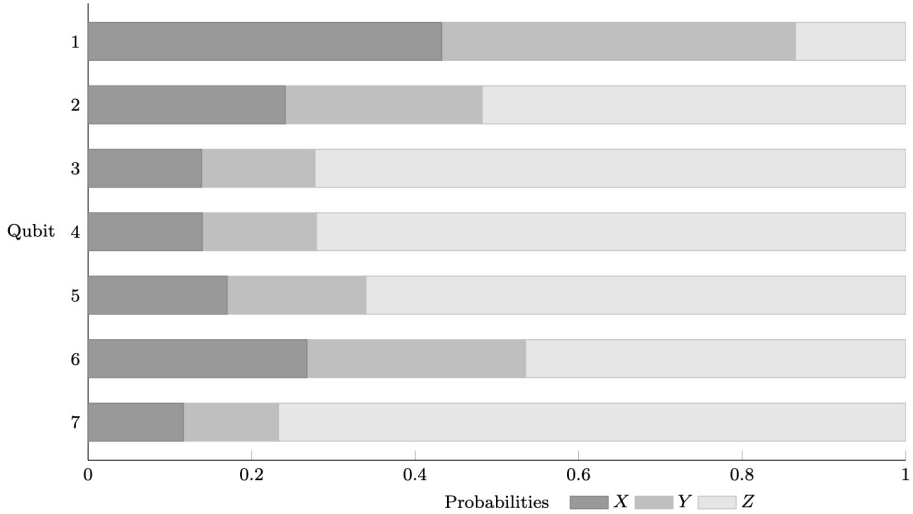


FIGURE 1. Probability distributions over the first 7 of 14 qubits for H_2O Hamiltonian using the Jordan-Wigner encoding. The probability distributions have been optimised according to Eq. (16).

Locally-biased classical shadows

Can knowledge of the coefficients *locally* improve random guessing à la classical shadows

$$\beta = \left(\prod_{i \in [n]} \beta_i \right) : \{X, Y, Z\}^n \to \mathbb{R}^+$$

$$\mathrm{Var}(\nu) \leq \mathbb{E}(\nu^2) = \sum_{P,Q} \alpha_P \alpha_Q g(P, Q, \beta) \mathrm{Tr}(PQ\rho)$$

Locally-biased classical shadows

Can knowledge of the coefficients *locally* improve random guessing à la classical shadows

$$\beta = \left(\prod_{i \in [n]} \beta_i \right) : \{X, Y, Z\}^n \to \mathbb{R}^+$$

$$\mathrm{Var}(\nu) \leq \mathbb{E}(\nu^2) = \sum_{P,Q} \alpha_P \alpha_Q g(P, Q, \beta) \mathrm{Tr}(PQ\rho)$$

$$\mathrm{cost}(\beta | \rho_{HF}) = \ldots$$

Locally-biased classical shadows

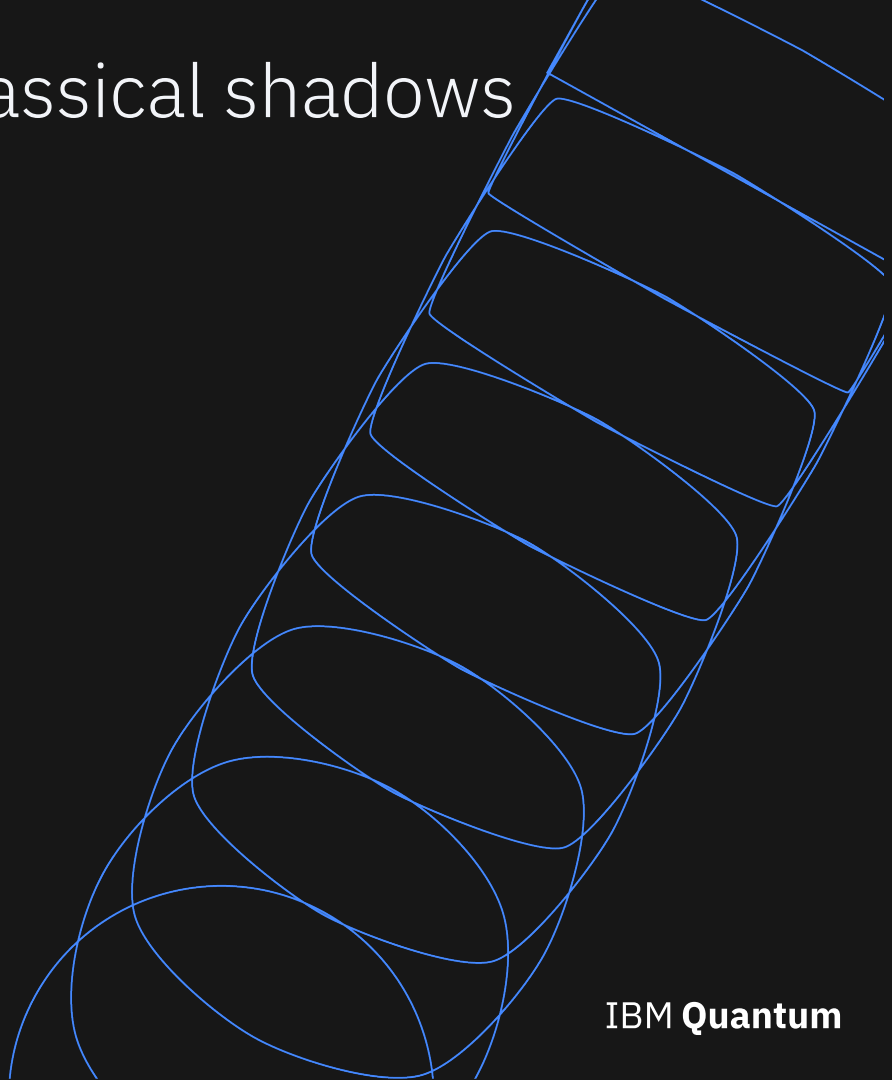
Can knowledge of the coefficients *locally* improve random guessing à la classical shadows

$$\beta = \left(\prod_{i \in [n]} \beta_i \right) : \{X, Y, Z\}^n \to \mathbb{R}^+$$

$$\mathrm{Var}(\nu) \leq \mathbb{E}(\nu^2) = \sum_{P,Q} \alpha_P \alpha_Q g(P, Q, \beta) \mathrm{Tr}(PQ\rho)$$

$$\mathrm{cost}_{\mathrm{diag}}(\beta) = \sum_P \alpha_P^2 \frac{1}{\prod_{i|P_i \neq I} \beta_i(P_i)}$$

Solution 3: Globally-biased classical shadows



IBM **Quantum**

Decision Diagrams for Quantum Measurements with Shallow Circuits

Stefan Hillmich,[1,*] Charles Hadfield,[2,†] Rudy Raymond,[3,4,‡] Antonio Mezzacapo,[2] and Robert Wille[1,5]

[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*

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 randomised measurements, which use decision diagrams to sample from probability distributions on measurement bases. This approach generalises previously known uniform and locally-biased randomised estimators. 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.

Decision diagrams

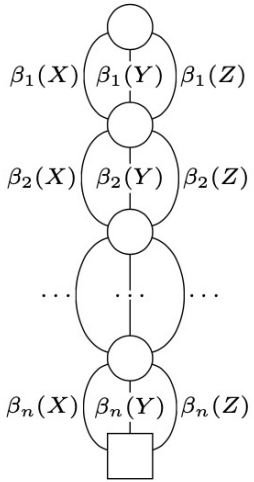


Figure 6: The Decision Diagram of LBCS

Decision diagrams

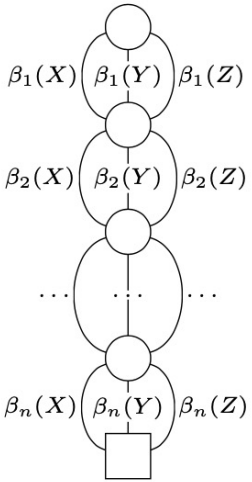


Figure 6: The Decision Diagram of LBCS

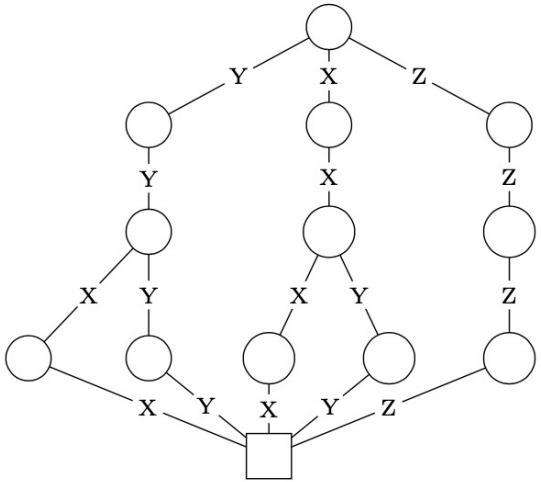


Figure 7: The unoptimised decision diagram of LDF-based Pauli Grouping of H_2 (4 qubits) in Jordan-Wigner encoding

Decision diagrams

Algorithm 2 Construction of a decision diagram (DD) from Hamiltonian H

Take absolute values of coefficients in H
Merge compatible terms to get reduced positive Pauli list $\mathcal{R}(H)$ ▷ Preprocessing
for Each term and coefficient in $\mathcal{R}(H)$ **do** ▷ Initialisation of DD
    Take existing path covering the longest prefix of term
    Create new edges for remaining Pauli operators up to the last
    Create edge to terminal with the last Pauli op and coefficient as edge weight
for Vertex in decision diagram in breadth-first order from terminal **do** ▷ Normalisation of DD
    Calculate sum of weights on out-going edges
    Divide weights on out-going edges by sum and multiply sum to in-coming edge weights
for Vertex in decision diagram in breadth-first order from terminal **do** ▷ Merge equivalent vertices in DD
    Calculate hash of vertex and if equivalent vertex exists, merge both
Remove identities in DD
    Replace "lonely" identity edges with virtual edges
    Remove identity edges where other edge with same source and target exists
    Merge targets of identity edges with target vertices of other edge
for Vertex in decision diagram in breadth-first order from terminal **do** ▷ Merge equivalent vertices in DD
    Calculate hash of vertex and if equivalent vertex exists, merge both
return decision diagram

Solution 4: Adaptive Pauli shadows

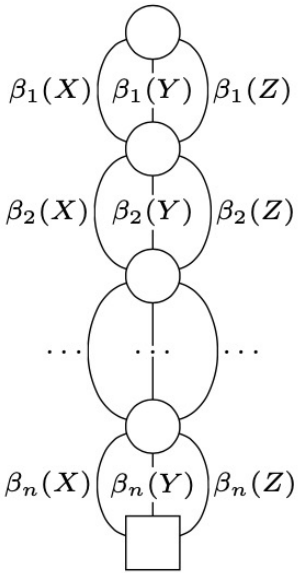
IBM **Quantum**

ADAPTIVE PAULI SHADOWS FOR ENERGY ESTIMATION

CHARLES HADFIELD

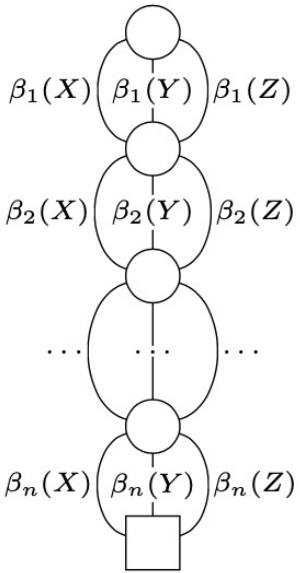
ABSTRACT. Locally-biased classical shadows allow rapid estimation of energies of quantum Hamiltonians. Recently, derandomised classical shadows have emerged claiming to be even more accurate. This accuracy comes at a cost of introducing classical computing resources into the energy estimation procedure. This present note shows, by adding a fraction of this classical computing resource to the locally-biased classical shadows setting, that the modified algorithm, termed *Adaptive Pauli Shadows* is state-of-the-art for energy estimation.

Adaptive Pauli shadows



$$\mathrm{cost}_{\mathrm{diag}}(\beta) = \sum_P \alpha_P^2 \frac{1}{\prod_{i|P_i \neq I} \beta_i(P_i)}$$

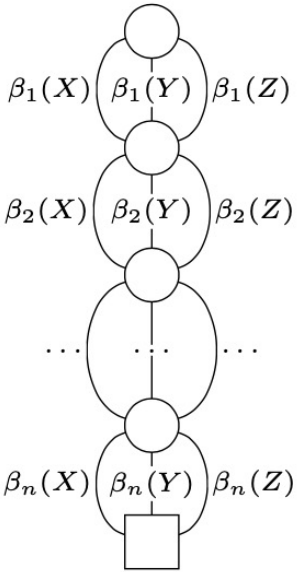
Adaptive Pauli shadows



Let’s choose each qubit’s probability *on-the-fly*.

$$\mathrm{cost}_{\mathrm{diag}}(\beta) = \sum_P \alpha_P^2 \frac{1}{\prod_{i|P_i \neq I} \beta_i(P_i)}$$

Adaptive Pauli shadows



Let's choose each qubit's probability *on-the-fly*.

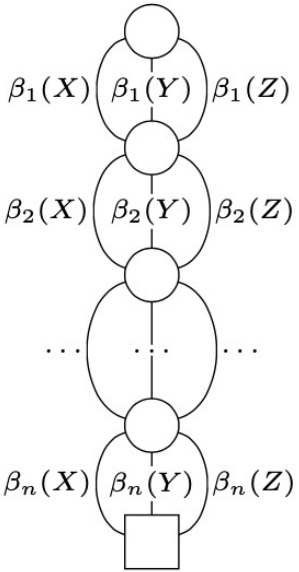
Let's start with the first qubit:

$$\text{cost}(\beta_1) = \frac{c_X}{\beta_1(X)} + \frac{c_Y}{\beta_1(Y)} + \frac{c_Z}{\beta_1(Z)}$$

Subject to beta being a probability distribution.

$$\text{cost}_{\text{diag}}(\beta) = \sum_P \alpha_P^2 \frac{1}{\prod_{i|P_i \neq I} \beta_i(P_i)}$$

Adaptive Pauli shadows



$$\text{cost}_{\text{diag}}(\beta) = \sum_P \alpha_P^2 \frac{1}{\prod_{i|P_i \neq I} \beta_i(P_i)}$$

Let's choose each qubit's probability *on-the-fly*.

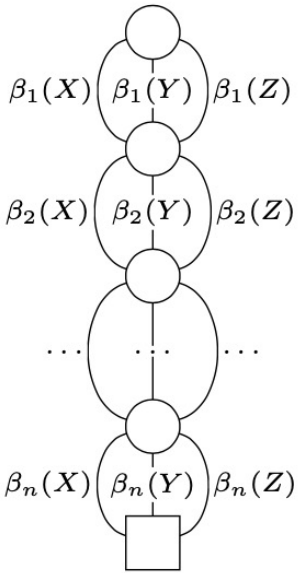
Let's start with the first qubit:

$$\text{cost}(\beta_1) = \frac{c_X}{\beta_1(X)} + \frac{c_Y}{\beta_1(Y)} + \frac{c_Z}{\beta_1(Z)}$$

Subject to beta being a probability distribution.

This cost function has an analytical solution!
Pick the basis B_1 from this distribution.

Adaptive Pauli shadows



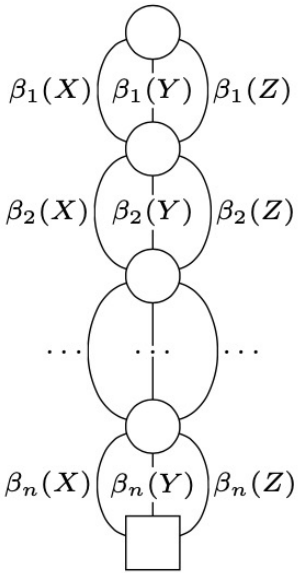
Let's choose each qubit's probability *on-the-fly*.

For i^{th} qubit, only look at Pauli terms for which it is still possible to provide an estimate upon eventual measurement

$$\mathrm{cost}(\beta_i) = \frac{c_X}{\beta_i(X)} + \frac{c_Y}{\beta_i(Y)} + \frac{c_Z}{\beta_i(Z)}$$

$$\mathrm{cost}_{\mathrm{diag}}(\beta) = \sum_P \alpha_P^2 \frac{1}{\prod_{i|P_i \neq I} \beta_i(P_i)}$$

Adaptive Pauli shadows



Let's choose each qubit's probability *on-the-fly*.

Algorithm 2 Choice of measurement basis for Adaptive Pauli Shadows

Randomly choose a bijection $i : [n] \to [n]$
for $j \in [n]$ **do**
  Set $\beta_{i(j)} : \mathcal{B} \to \mathbb{R}^+$ by solving the optimisation problem in Eq. (2)
  Choose $B_{i(j)}$ randomly according to distribution $\beta_{i(j)}$
return $B = \otimes_{i\in[n]} B_i$.

$$\mathrm{cost}_{\mathrm{diag}}(\beta) = \sum_P \alpha_P^2 \frac{1}{\prod_{i|P_i\neq I} \beta_i(P_i)}$$

Time to run VQE for some molecules

		Qiskit	LBCS	GBCS	Caltech	APS
LiH	JW	55	15	10	1	1.6
	P	85	30	15	1	2.5
	BK	75	70	15	1.6	5
BeH_2	JW	140	70	30	3.6	3.6
	P	240	130	40	8.1	3.6
	BK	200	240	60	3.6	3.6
H_2O	JW	1000	260	300	14	12
	P	2700	430	430	50	12
	BK	2100	1400	530	40	10
NH_3	JW	900	350		32	16
	P	2600	630		44	19
	BK	2150	380		14	12