

Logical shadow tomography: Efficient estimation of error-mitigated observables

Hong-Ye Hu,^{1,2,3,*} Ryan LaRose,^{4,2,*} Yi-Zhuang You,¹ Eleanor Rieffel,² and Zhihui Wang^{2,3,†}

¹*Department of Physics, University of California San Diego, La Jolla, CA 92093, USA*

²*Quantum Artificial Intelligence Laboratory (QuAIL),*

NASA Ames Research Center, Moffett Field, CA, 94035, USA

³*USRA Research Institute for Advanced Computer Science (RIACS), Mountain View, CA, 94043, USA*

⁴*Department of Computational Mathematics, Science, and Engineering, Michigan State University*

We introduce a technique to estimate error-mitigated expectation values on noisy quantum computers. Our technique performs shadow tomography on a logical state to produce a memory-efficient classical reconstruction of the noisy density matrix. Using efficient classical post-processing, one can mitigate errors by projecting into the codespace as in subspace expansion and taking powers of the density matrix as in virtual distillation. Relative to subspace expansion which requires $\Omega(2^{(n-1)k})$ samples to estimate a Pauli observable with an $[[n, k]]$ stabilizer code, our technique requires only $\Theta(2^k)$ samples. Relative to virtual distillation, our technique can compute powers of the density matrix without implementing additional copies of quantum states the quantum computer. We present numerical results using logical states encoded with up to sixty physical qubits and show fast convergence to error-free expectation values with only 10^5 samples under 1% depolarizing noise.

I. INTRODUCTION

Recent advances in experimental quantum information processing have enabled complex simulations of interesting physical systems [1–4]. A common and crucial element of each experiment is mitigating the effect of errors, a task which is generally referred to as quantum error mitigation (QEM) [5].

A relatively large number of QEM techniques have been proposed in recent literature [6–10]. Some authors have defined common frameworks which encapsulate one or more of these techniques [11, 12]. At its core, any QEM technique uses additional quantum resources (qubits, gates, and/or samples) in a clever way to approximate what would happen in an ideal device.

In this work we present a conceptually simple and practical QEM technique. The only step which happens on a quantum computer is repeatedly sampling from an encoded state to perform shadow tomography [13, 14]. After this, one can use classical post-processing to project out errors as in subspace expansion [8] and calculate powers of the density matrix as in virtual distillation [10]. In addition to enabling both of these methods, our technique requires significantly fewer resources and can be practical for relatively large systems.

II. LOGICAL SHADOW TOMOGRAPHY

A. Setup

We wish to estimate the expectation value of an observable Γ with respect to a pure state $|\psi\rangle$:

$$\langle\Gamma\rangle := \langle\psi|\Gamma|\psi\rangle = \text{Tr}[\rho|\psi\rangle\langle\psi|\Gamma]. \quad (1)$$

If we use a quantum computer, the resulting state is noisy and we instead evaluate

$$\langle\Gamma\rangle_{\text{noisy}} := \text{Tr}[\mathcal{E}(|\psi\rangle\langle\psi|)\Gamma] \quad (2)$$

where $\mathcal{E}$ is an unknown quantum channel which depends on the quantum computer being used. The goal of quantum error mitigation is to compute a quantity $\langle\Gamma\rangle_{\text{QEM}}$ such that

$$|\langle\Gamma\rangle_{\text{QEM}} - \langle\Gamma\rangle| < |\langle\Gamma\rangle_{\text{noisy}} - \langle\Gamma\rangle|. \quad (3)$$

B. Procedure

Let us now introduce our technique which we will refer to as logical shadow tomography (LST). LST consists of the following steps¹ (see Fig. 1 for a graphical overview):

1. Given a k -qubit state $|\psi\rangle$ and an $[[n, k, d]]$ stabilizer code $\mathcal{S} = \langle S_1, \dots, S_{n-k} \rangle$, construct an n -qubit circuit that prepares the logical state $\bar{\rho} \equiv |\bar{\psi}\rangle\langle\bar{\psi}|$.
2. Append a randomly sampled unitary U from unitary ensemble $\mathcal{E}_U$ to the logical state preparation circuit and measure all qubits in the computational basis to obtain a bitstring $b \in \{0, 1\}^n$. Store U and b .
3. Repeat step 2 N times to obtain $U_1, \dots, U_N$ and $b_1, \dots, b_N$.

One then uses a classical computer to form the classical shadow of $\mathcal{E}(\bar{\rho})$

$$S(\mathcal{E}(\bar{\rho}), N) := \left\{ \hat{\rho}_i := \mathcal{M}^{-1} \left(U_i^\dagger |b_i\rangle\langle b_i| U_i \right) \right\}_{i=1}^N \quad (4)$$

* Equal contributions.

† Corresponding author: zhihui.wang@nasa.gov.

¹ Steps 2 and 3 can be aptly summarized by “perform shadow tomography [14] on the logical state,” whence *logical shadow tomography*. We explain these steps in detail for completeness.

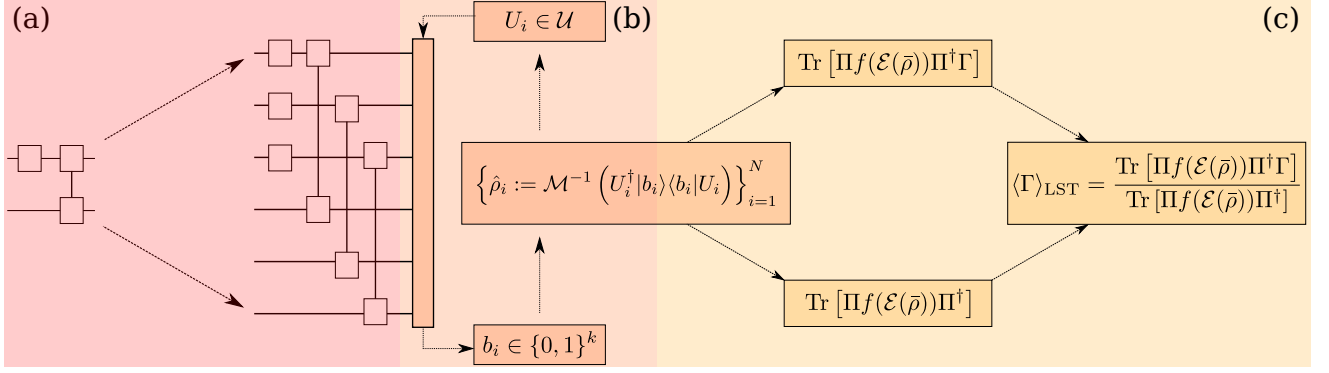


FIG. 1. Schematic workflow of the main steps of logical shadow tomography (LST). (a) Given an n -qubit circuit preparing the state $|\psi\rangle$, an $[[n, k, d]]$ stabilizer code is used to construct a k -qubit circuit that prepares the logical state $|\bar{\psi}\rangle$. (b) For a unitary ensemble $\mathcal{U}$ and an integer N , an element U_i is randomly sampled from $\mathcal{U}$ and appended to the circuit. All qubits are measured in the computational basis to get a bitstring $b_i \in \{0, 1\}^k$. This process is repeated N times to form the classical shadow $S(\mathcal{E}(\bar{\rho}), N) := \left\{ \hat{\rho}_i := \mathcal{M}^{-1} \left(U_i^\dagger |b_i\rangle \langle b_i| U_i \right) \right\}_{i=1}^N$. The map $\mathcal{M}^{-1}$ is determined by $\mathcal{U}$, for example $\mathcal{M}^{-1}(A) = 2^n A - I$ if $\mathcal{U}$ is a n -qubit global Clifford ensemble. The channel $\mathcal{E}$ characterizes the noisy quantum computer from which samples are drawn. (c) Classical postprocessing is used to estimate two quantities from the shadow $S(\mathcal{E}(\bar{\rho}), N)$; the final LST output is the ratio of the two. The projector Π is defined by the chosen stabilizer code, f is an arbitrary function, and Γ is the observable we wish to compute the expectation value of decomposed in the Pauli basis.

where the map $\mathcal{M}^{-1}$ depends on the unitary ensemble $\mathcal{U}$. There are a few proposals for the random unitary ensembles, for example using random Clifford circuits and using chaotic dynamics[HYH:ref]. In our paper, with out loss of generality, we will consider random unitary ensemble consisting of Clifford gates.

...TBC...

for example $\mathcal{M}^{-1}(A) = (2^n + 1)A - I$ if $\mathcal{U}$ is an n -qubit Clifford ensemble and $\mathcal{M}^{-1}(A) = A$ if $\mathcal{U}$ is an ensemble of n -fold tensor products of single-qubit Cliffords [14]. In either case, note that each $U_i^\dagger |b_i\rangle$ is a stabilizer state and so the classical shadow can be stored efficiently in classical memory. The final output of LST is an estimate of the quantity

$$\langle \Gamma \rangle_{\text{LST}} = \frac{\text{Tr} [\Pi f(\mathcal{E}(\bar{\rho})) \Pi^\dagger \Gamma]}{\text{Tr} [\Pi f(\mathcal{E}(\bar{\rho})) \Pi^\dagger]} \quad (5)$$

where f is an arbitrary matrix function, for example $f(\mathcal{E}(\bar{\rho})) = \mathcal{E}(\bar{\rho})^2$, and Π is the projection onto the codespace

$$\Pi := \prod_{i=1}^{n-k} (I + S_i)/2 \quad (6)$$

The LST output (5) is computed via a median of means estimate as in [14]. For simplicity we present linear case $f(\mathcal{E}(\bar{\rho})) = \mathcal{E}(\bar{\rho})$. The steps are as follows:

4. Split the shadow (4) into K equally-sized parts and set

$$\hat{\rho}_{(k)} = \frac{1}{N/K} \sum_{i=kN/K+1}^{kN/K} \hat{\rho}_i \quad (7)$$

where for simplicity we assumed N/K is integral.

5. Estimate the numerator of (5) via

$$\text{Tr}[Pf(\bar{\rho})P^\dagger \Gamma] \leftarrow \text{median} \{ \text{Tr}[\hat{\rho}_{(1)} P \Gamma], \dots, \text{Tr}[\hat{\rho}_{(K)} P \Gamma] \}. \quad (8)$$

6. Estimate the denominator of (5) via

$$\text{Tr}[Pf(\bar{\rho})P^\dagger] \leftarrow \text{median} \{ \text{Tr}[\hat{\rho}_{(1)} P], \dots, \text{Tr}[\hat{\rho}_{(K)} P] \}. \quad (9)$$

7. Output the ratio $\text{Tr}[Pf(\bar{\rho})P^\dagger \Gamma] / \text{Tr}[Pf(\bar{\rho})P^\dagger]$.

C. Analysis

1. Error mitigation capability

Let $\mathcal{S}$ be the stabilizer code used in LST. For any correctable error E , there exists a stabilizer generator S such that $SE = -ES$ and so

$$\Pi E |\psi\rangle \propto (I + S)E |\psi\rangle = |\psi\rangle - |\psi\rangle = 0. \quad (10)$$

Analogously, if no error has occurred then $|\psi\rangle$ is a code-word and

$$\Pi |\psi\rangle = |\psi\rangle. \quad (11)$$

Thus the projector Π discards results in which correctable errors have occurred [8]. The set of correctable errors is determined by the chosen code.

Suppose now that the noisy state $\mathcal{E}(\rho)$ has spectral decomposition

$$\mathcal{E}(\rho) = p_0 |\psi_0\rangle \langle \psi_0| + p_1 |\psi_1\rangle \langle \psi_1| + \dots + p_n |\psi_n\rangle \langle \psi_n| \quad (12)$$

where $|\psi_0\rangle \equiv |\psi\rangle$ is the noiseless state, $p_0 > \dots > p_n$, and $\langle\psi_i|\psi_j\rangle = \delta_{ij}$. In this case, for any observable O and positive integer m we have [10, 11]

$$\frac{\text{Tr}[\mathcal{E}(\rho)^m O]}{\text{Tr}[\mathcal{E}(\rho)^m]} = \frac{p_0^m \langle\psi_0|O|\psi_0\rangle + \sum_i p_i^m \langle\psi_i|O|\psi_i\rangle}{p_0^m + \sum_i p_i^m} \quad (13)$$

$$= \langle\psi|O|\psi\rangle [1 + O([p_1/p_0]^m)]. \quad (14)$$

Thus, computing the expectation value with the m th power of the state suppresses errors exponentially in m , a phenomenon which can also be interpreted as artificially cooling the system [15]. Here we allow for an arbitrary function f acting on the noisy state via its Taylor expansion

$$f(x) = a_0 x + \dots + a_m x^m \quad (15)$$

Including sums of powers up to m (instead of just the highest power m) was shown to improve results of numerical experiments in [11].

2. Quantum resources

LST requires additional qubits and gates to encode the logical state, the exact number of which depends on the chosen code. Independent of the code, an element from a unitary ensemble $\mathcal{U}$ is appended to the logical circuit before measuring all qubits in the computational basis. If $\mathcal{U}$ is chosen to be a tensor-product Pauli ensemble, the added circuit depth is $O(1)$ [14]. If $\mathcal{U}$ is chosen to be a global Clifford ensemble, the added circuit depth is $O(n)$ [16].

The other quantum resource is measurements to produce samples (bitstrings). From the result of classical shadow tomography [14], we have the following theorem which dictates the sample complexity of LST.

Theorem 1 *Let ρ be a quantum state and $\langle O \rangle := \text{Tr}[\rho O]$ be the expectation value we wish to estimate. Then, classical shadow tomography can produce an estimate $\langle \tilde{O} \rangle$ such that*

$$\Pr\left(\left|\langle \tilde{O} \rangle - \langle O \rangle\right| \geq \delta\right) \leq \epsilon \quad (16)$$

using $\mathcal{O}(4/\epsilon\delta^2)$ independent copies of ρ .

The result (16) applies to the numerator ($O = P^\dagger \Gamma P$) and denominator ($O = P^\dagger P$) of the LST estimate (5) separately. We emphasize that the number of samples does not depend on the number of qubits and so does not depend on the chosen $[[n, k, d]]$ code.

3. Classical resources

After sampling, we have the classical shadow (4) consisting of N stabilizer states, the observable Γ , and the

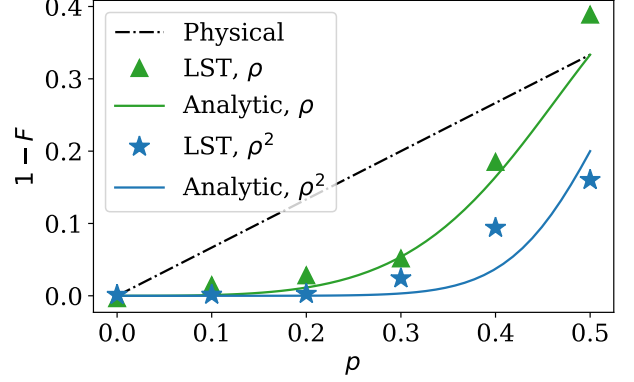


FIG. 2. LST with the $[[5, 1, 3]]$ code. Here, $|\psi\rangle$ is taken to be the ground state of the code and the observable is taken to be $\Gamma = |\psi\rangle\langle\psi|$. The dashed black line shows the physical infidelity, i.e., the noisy expectation value. The green points show the LST estimate (5) with $f(\mathcal{E}(\bar{\rho})) = \mathcal{E}(\bar{\rho})$ and the blue points show the LST estimate with $f(\mathcal{E}(\bar{\rho})) = \mathcal{E}(\bar{\rho})^2$. Solid lines show the corresponding analytic values, i.e., the values computed exactly without sampling.

projector Π . We need to estimate the numerator and denominator of (5), both of which can be written

$$\text{Tr}[f(\mathcal{E}(\bar{\rho}))O] \quad (17)$$

where $O = \Gamma\Pi$ for the numerator and $O = \Pi$ for the denominator.

Let m be the highest degree of the Taylor expansion of f . As in [], we can evaluate the expectation of this term via

$$\mathbb{E} \text{Tr}[\hat{\rho}_{i_1} \otimes \dots \otimes \hat{\rho}_{i_m} S O] \quad (18)$$

where S is a permutation of the indices $i_1, \dots, i_m$. A single snapshot $\hat{\rho}_i$ requires $O(2n)$ classical memory to store, so the argument inside the trace in (18) requires $O((2n)^m)$ classical memory. Evaluating the expectation requires m summations over N snapshots and so has time $O(N^m)$.

III. NUMERICAL RESULTS

In Sec. IIC1 we discussed the error mitigation capabilities of LST. We now show two numerical experiments to demonstrate the performance in practice.

In both experiments we consider a single-qubit state $|\psi\rangle$ and take $\Gamma = |\psi\rangle\langle\psi|$ as the observable, i.e., we estimate the overlap of the LST estimate with the true noiseless state.

A. Pseudo-threshold with the $[[5, 1, 3]]$ code

We first consider a single-qubit depolarizing noise model for various noise strengths $0 \leq p \leq 1/2$ and per-

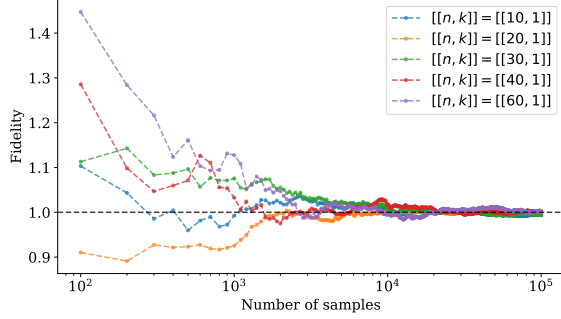


FIG. 3. LST estimated fidelity vs. number of samples ranging from 10^2 - 10^5 with various code sizes. The noiseless fidelity value of 1.0 is shown with the dashed black line. For all code sizes up to $n = 60$ physical qubits, the LST estimate converges to the true noiseless value. Codes used are the minimum distance constructions from [1].

form LST using the $[[5, 1, 3]]$ code

$$\mathcal{S} = \langle XZZXI, IXZZX, XIXZZ, ZXIXZ \rangle. \quad (19)$$

We evaluate (5) with $f(x) = x$ and $f(x) = x^2$. The results are shown in Fig. 2. Here, the dashed black line shows the infidelity without any error mitigation (the “physical” curve), and the solid colored lines/points show the infidelity using LST. We see that the LST estimates have lower infidelity than the physical curve, showing that errors are indeed being mitigated. LST with $f(x) = x^2$ outperforms LST with $f(x) = x$ showing that the combination of projecting into the codespace and virtually cooling the system outperforms only projecting into the codespace. We expect higher powers $f(x) = x^m$ with $m > 2$ to mitigate errors even further as in [10]. In both cases, the LST estimates evaluated by sampling are shown with points, and the solid colored lines show the exact (“analytic”) value without sampling.

B. Convergence vs. code size

We now consider $[[n, 1, d]]$ codes and vary the number of physical qubits ranges from $n = 10$ to $n = 60$. The LST estimated fidelity (using $f(x) = x$ for all code sizes) vs. number of samples is shown in Fig. 3. Using a relatively small number of samples (at most 10^5), all LST values converge to the noiseless fidelity. Note that the direct implementation of subspace expansion with the full projection Π as used here would require at least 2^{59} sam-

ples, a number infeasible to implement in any practical experiment.

IV. DISCUSSION

We have presented a procedure for estimating error-mitigated observables on noisy quantum computers. Our procedure is flexible enough to be performed on virtually any quantum computer: the only additional quantum resources needed are qubits and gates for encoding the logical state. After sampling from the logical state, a classical computer processes the obtained classical shadow to return the error-mitigated expectation value. This classical processing is efficient in the complexity-theoretic sense. Further, it can be easily parallelized for practical efficiency in real-world experiments.

With respect to error mitigation, our procedure provides an experimentally simple procedure to carry out proposed error mitigation techniques at scale. In particular, we have demonstrated subspace expansion with up to $n = 60$ physical qubits encoding a single logical qubit, i.e., a stabilizer group with 2^{59} elements, an experiment which would be practically infeasible with the direct or stochastic sampling schemes of [8, 11]. We have also demonstrated virtual distillation [10] without expensive swap operations to compute powers of the density matrix. Rather, our procedure uses the same quantum circuit to evaluate the error-mitigated expectation with any function $f(\mathcal{E}(\rho))$ of the noisy state; the only difference is in classical postprocessing. Further, beyond making both of these techniques significantly more practical to implement, our procedure enables them to be composed with one another. We have shown numerically the composition of both techniques results in further reduction of errors.

Our procedure can also be seen as an application of shadow tomography [14]. Beyond the multiple applications presented in the original work, additional applications have emerged in recent literature, for example performing process tomography [17]. This work constitutes another application, namely error mitigation.

We are optimistic our procedure will be useful on current and near-term quantum computers for a large variety of experiments on relatively large systems. We plan to investigate this in future work.

Acknowledgements We are grateful for support from the NASA Ames Research Center and from the DARPA ONISQ program under interagency agreement IAA 8839, Annex 114. Z.W. is supported by USRA Nasa Academic Mission Service (NNA16BD14C). H.Y.H. is supported by the USRA Feynman Quantum Academy funded by the NAMS R&D Student Program. H.Y.H. and Y.Z.Y. are supported by a UC Hellman Fellowship. R.L. acknowledges funding from a NASA Space Technology Graduate Research Fellowship.

-
- [1] Frank Arute, Kunal Arya, Ryan Babbush, Dave Bacon, Joseph C. Bardin, Rami Barends, Sergio Boixo, Michael Broughton, Bob B. Buckley, David A. Buell, Brian Burkett, Nicholas Bushnell, Yu Chen, Zijun Chen, Ben Chiaro, Roberto Collins, William Courtney, Sean Demura, Andrew Dunsworth, Edward Farhi, Austin Fowler, Brooks Foxen, Craig Gidney, Marissa Giustina, Rob Graff, Steve Habegger, Matthew P. Harrigan, Alan Ho, Sabrina Hong, Trent Huang, L. B. Ioffe, Sergei V. Isakov, Evan Jeffrey, Zhang Jiang, Cody Jones, Dvir Kafri, Kostyantyn Kechedzhi, Julian Kelly, Seon Kim, Paul V. Klimov, Alexander N. Korotkov, Fedor Kostritsa, David Landhuis, Pavel Laptev, Mike Lindmark, Martin Leib, Erik Lucero, Orion Martin, John M. Martinis, Jarrod R. McClean, Matt McEwen, Anthony Megrant, Xiao Mi, Masoud Mohseni, Wojciech Mroczkiewicz, Josh Mutus, Ofer Naaman, Matthew Neeley, Charles Neill, Florian Neukart, Hartmut Neven, Murphy Yuezhen Niu, Thomas E. O'Brien, Bryan O'Gorman, Eric Ostby, Andre Petukhov, Harald Putterman, Chris Quintana, Pedram Roushan, Nicholas C. Rubin, Daniel Sank, Kevin J. Satzinger, Andrea Skolik, Vadim Smelyanskiy, Doug Strain, Michael Streif, Kevin J. Sung, Marco Szalay, Amit Vainsencher, Theodore White, Z. Jamie Yao, Ping Yeh, Adam Zalcman, and Leo Zhou. Quantum Approximate Optimization of Non-Planar Graph Problems on a Planar Superconducting Processor. *arXiv:2004.04197 [quant-ph]*, April 2020. arXiv: 2004.04197.
- [2] Frank Arute, Kunal Arya, Ryan Babbush, Dave Bacon, Joseph C. Bardin, Rami Barends, Andreas Bengtsson, Sergio Boixo, Michael Broughton, Bob B. Buckley, David A. Buell, Brian Burkett, Nicholas Bushnell, Yu Chen, Zijun Chen, Yu-An Chen, Ben Chiaro, Roberto Collins, Stephen J. Cotton, William Courtney, Sean Demura, Alan Derk, Andrew Dunsworth, Daniel Eppens, Thomas Eckl, Catherine Erickson, Edward Farhi, Austin Fowler, Brooks Foxen, Craig Gidney, Marissa Giustina, Rob Graff, Jonathan A. Gross, Steve Habegger, Matthew P. Harrigan, Alan Ho, Sabrina Hong, Trent Huang, William Huggins, Lev B. Ioffe, Sergei V. Isakov, Evan Jeffrey, Zhang Jiang, Cody Jones, Dvir Kafri, Kostyantyn Kechedzhi, Julian Kelly, Seon Kim, Paul V. Klimov, Alexander N. Korotkov, Fedor Kostritsa, David Landhuis, Pavel Laptev, Mike Lindmark, Erik Lucero, Michael Marthaler, Orion Martin, John M. Martinis, Anika Maruszyk, Sam McArdle, Jarrod R. McClean, Trevor McCourt, Matt McEwen, Anthony Megrant, Carlos Mejuto-Zaera, Xiao Mi, Masoud Mohseni, Wojciech Mroczkiewicz, Josh Mutus, Ofer Naaman, Matthew Neeley, Charles Neill, Hartmut Neven, Michael Newman, Murphy Yuezhen Niu, Thomas E. O'Brien, Eric Ostby, Bálint Pató, Andre Petukhov, Harald Putterman, Chris Quintana, Jan-Michael Reiner, Pedram Roushan, Nicholas C. Rubin, Daniel Sank, Kevin J. Satzinger, Vadim Smelyanskiy, Doug Strain, Kevin J. Sung, Peter Schmitteckert, Marco Szalay, Norm M. Tubman, Amit Vainsencher, Theodore White, Nicolas Vogt, Z. Jamie Yao, Ping Yeh, Adam Zalcman, and Sebastian Zanker. Observation of separated dynamics of charge and spin in the Fermi-Hubbard model. *arXiv:2010.07965 [quant-ph]*, October 2020. arXiv: 2010.07965.
- [3] Xiao Mi, Pedram Roushan, Chris Quintana, Salvatore Mandra, Jeffrey Marshall, Charles Neill, Frank Arute, Kunal Arya, Juan Atalaya, Ryan Babbush, Joseph C. Bardin, Rami Barends, Andreas Bengtsson, Sergio Boixo, Alexandre Bourassa, Michael Broughton, Bob B. Buckley, David A. Buell, Brian Burkett, Nicholas Bushnell, Zijun Chen, Benjamin Chiaro, Roberto Collins, William Courtney, Sean Demura, Alan R. Derk, Andrew Dunsworth, Daniel Eppens, Catherine Erickson, Edward Farhi, Austin G. Fowler, Brooks Foxen, Craig Gidney, Marissa Giustina, Jonathan A. Gross, Matthew P. Harrigan, Sean D. Harrington, Jeremy Hilton, Alan Ho, Sabrina Hong, Trent Huang, William J. Huggins, L. B. Ioffe, Sergei V. Isakov, Evan Jeffrey, Zhang Jiang, Cody Jones, Dvir Kafri, Julian Kelly, Seon Kim, Alexei Kitaev, Paul V. Klimov, Alexander N. Korotkov, Fedor Kostritsa, David Landhuis, Pavel Laptev, Erik Lucero, Orion Martin, Jarrod R. McClean, Trevor McCourt, Matt McEwen, Anthony Megrant, Kevin C. Miao, Masoud Mohseni, Wojciech Mroczkiewicz, Josh Mutus, Ofer Naaman, Matthew Neeley, Michael Newman, Murphy Yuezhen Niu, Thomas E. O'Brien, Alex Opremcak, Eric Ostby, Balint Pato, Andre Petukhov, Nicholas Redd, Nicholas C. Rubin, Daniel Sank, Kevin J. Satzinger, Vladimir Shvarts, Doug Strain, Marco Szalay, Matthew D. Trevithick, Benjamin Villalonga, Theodore White, Z. Jamie Yao, Ping Yeh, Adam Zalcman, Hartmut Neven, Igor Aleiner, Kostyantyn Kechedzhi, Vadim Smelyanskiy, and Yu Chen. Information Scrambling in Computationally Complex Quantum Circuits. *Science*, page eabg5029, October 2021. arXiv: 2101.08870.
- [4] M. S. Blok, V. V. Ramasesh, T. Schuster, K. O'Brien, J. M. Kreikebaum, D. Dahlen, A. Morvan, B. Yoshida, N. Y. Yao, and I. Siddiqi. Quantum Information Scrambling in a Superconducting Qutrit Processor. *Physical Review X*, 11(2):021010, April 2021. arXiv: 2003.03307.
- [5] Ryuji Takagi, Suguru Endo, Shintaro Minagawa, and Mile Gu. Fundamental limits of quantum error mitigation. *arXiv:2109.04457 [quant-ph]*, October 2021. arXiv: 2109.04457.
- [6] Ying Li and Simon C. Benjamin. Efficient Variational Quantum Simulator Incorporating Active Error Minimization. *Physical Review X*, 7(2):021050, June 2017.
- [7] Kristan Temme, Sergey Bravyi, and Jay M. Gambetta. Error mitigation for short-depth quantum circuits. *Physical Review Letters*, 119(18):180509, November 2017. arXiv: 1612.02058.
- [8] Jarrod R. McClean, Zhang Jiang, Nicholas C. Rubin, Ryan Babbush, and Hartmut Neven. Decoding quantum errors with subspace expansions. *arXiv:1903.05786 [physics, physics:quant-ph]*, March 2019. arXiv: 1903.05786.
- [9] Piotr Czarnik, Andrew Arrasmith, Patrick J. Coles, and Lukasz Cincio. Error mitigation with Clifford quantum-circuit data. *arXiv:2005.10189 [quant-ph]*, February 2021. arXiv: 2005.10189.
- [10] William J. Huggins, Sam McArdle, Thomas E. O'Brien, Joonho Lee, Nicholas C. Rubin, Sergio Boixo, K. Birgitta Whaley, Ryan Babbush, and Jarrod R. McClean. Virtual Distillation for Quantum Error Mitigation. *arXiv:2011.07064 [quant-ph]*, August 2021. arXiv: 2011.07064.

- 2011.07064.
- [11] Nobuyuki Yoshioka, Hideaki Hakoshima, Yuichiro Matsuzaki, Yuuki Tokunaga, Yasunari Suzuki, and Suguru Endo. Generalized quantum subspace expansion. *arXiv:2107.02611 [quant-ph]*, July 2021. arXiv: 2107.02611.
- [12] Angus Lowe, Max Hunter Gordon, Piotr Czarnik, Andrew Arrasmith, Patrick J. Coles, and Lukasz Cincio. Unified approach to data-driven quantum error mitigation. *arXiv:2011.01157 [quant-ph]*, November 2020. arXiv: 2011.01157.
- [13] Scott Aaronson. Shadow Tomography of Quantum States. *arXiv:1711.01053 [quant-ph]*, November 2018. arXiv: 1711.01053.
- [14] Hsin-Yuan Huang, Richard Kueng, and John Preskill. Predicting many properties of a quantum system from very few measurements. *Nature Physics*, 16(10):1050–1057, October 2020. Bandiera_abtest: a Cg_type: Nature Research Journals Number: 10 Primary_atype: Research Publisher: Nature Publishing Group Subject_term: Information theory and computation; Mathematics and computing; Quantum information; Quantum physics; Theoretical physics Subject_term_id: information-theory-and-computation; mathematics-and-computing; quantum-information; quantum-physics; theoretical-physics.
- [15] Jordan Cotler, Soonwon Choi, Alexander Lukin, Hrant Gharibyan, Tarun Grover, M. Eric Tai, Matthew Rispoli, Robert Schittko, Philipp M. Preiss, Adam M. Kaufman, Markus Greiner, Hannes Pichler, and Patrick Hayden. Quantum Virtual Cooling. *Physical Review X*, 9(3):031013, July 2019. arXiv: 1812.02175.
- [16] Hong-Ye Hu and Yi-Zhuang You. Hamiltonian-Driven Shadow Tomography of Quantum States. *arXiv:2102.10132 [quant-ph]*, March 2021. arXiv: 2102.10132.
- [17] Ryan Levy, Di Luo, and Bryan K. Clark. Classical Shadows for Quantum Process Tomography on Near-term Quantum Computers. *arXiv:2110.02965 [cond-mat, physics:physics, physics:quant-ph]*, October 2021. arXiv: 2110.02965.

Appendix A: Details on shadow tomography and proof of results

1. Mean and variance of a ratio of two random variables

Consider random variables P and Q and let $G = g(P, Q) = P/Q$. In general, there is no close form expression for $\mathbb{E}[G(P, Q)]$, and $\text{Var}[G(P, Q)]$. Here, we find approximations for the mean and variance using Taylor expansions of $g(P, Q)$.

The approximation for the mean value is

$$\begin{aligned}\mathbb{E}[g(P, Q)] &= \mathbb{E}[g(\mu_P, \mu_Q) + g'_P(\mu_P, \mu_Q)(P - \mu_P) + g'_Q(\mu_P, \mu_Q)(Q - \mu_Q) + R] \\ &\approx \mathbb{E}[g(\mu_P, \mu_Q)] + g'_P(\mu_P, \mu_Q)\mathbb{E}[(P - \mu_P)] + g'_Q(\mu_P, \mu_Q)\mathbb{E}[(Q - \mu_Q)] \\ &= g(\mu_P, \mu_Q),\end{aligned}\tag{A1}$$

where R is the higher order reminders of the Taylor expansion. For keeping the Taylor expansion to the first order, we ignore higher order reminders.

For the variance, we have

$$\begin{aligned}\text{Var}[g(P, Q)] &= \mathbb{E}\{[g(P, Q) - \mathbb{E}(g(P, Q))]^2\} \\ &\approx \mathbb{E}\{[g(\mu_P, \mu_Q) + g'_P(\mu_P, \mu_Q)(P - \mu_P) + g'_Q(\mu_P, \mu_Q)(Q - \mu_Q)]^2\} \\ &\approx \mathbb{E}\{[g'_P(\mu_P, \mu_Q)(P - \mu_P) + g'_Q(\mu_P, \mu_Q)(Q - \mu_Q)]^2\} \\ &= g_P'^2(\mu_P, \mu_Q)\text{Var}(P) + g_Q'^2(\mu_P, \mu_Q)\text{Var}(Q) + 2g'_P(\mu_P, \mu_Q)g'_Q(\mu_P, \mu_Q)\text{Cov}(P, Q).\end{aligned}\tag{A2}$$

For our case, $g(P, Q) = P/Q$, therefore $g'_P(\mu_P, \mu_Q) = 1/\mu_Q$, $g'_Q(\mu_P, \mu_Q) = -\mu_P/\mu_Q^2$, and

$$\text{Var}(P/Q) \approx \left(\frac{\mu_P}{\mu_Q}\right)^2 \left[\frac{\text{Var}(P)}{\mu_P^2} + \frac{\text{Var}(Q)}{\mu_Q^2} - 2\frac{\text{Cov}(P, Q)}{\mu_P\mu_Q} \right]\tag{A3}$$

2. Proof of sample complexities

Suppose we want to predict a linear properties of the underlying quantum state,

$$o = \text{Tr}(\rho O).\tag{A4}$$

We simply replace the unknown quantum state ρ with the classical shadows $\hat{\rho} = \mathcal{M}^{-1}(\hat{\sigma})$. This yields a stochastic number $\hat{o} = \text{Tr}(\hat{\rho}O)$, and it will converge to correct answer with sufficient amount of classical shadows,

$$\mathbb{E}\hat{o} = \text{Tr}(\rho O). \quad (\text{A5})$$

In practice, the expectation $\mathbb{E}\hat{o}_i$ is replaced by a sample mean estimator, $o_{\text{avg}} = \frac{1}{M} \sum_{i=1}^M \hat{o}_i = \frac{1}{M} \sum_{i=1}^M \text{Tr}(O\hat{\rho}_i)$. Based on Chebyshev's inequality, the probability of the estimation o_{avg} to deviate from its expectation value o is bounded by its variance $\text{Var}(o_{\text{avg}})$ as $\Pr(|o_{\text{avg}} - o| \geq \delta) \leq \text{Var}(o_{\text{avg}})/\delta^2$. To control the deviation within a desired statistical accuracy ϵ , we require $\text{Var}(o_{\text{avg}})/\delta^2 = \text{Var}(\hat{o})/(M\delta^2) \leq \epsilon$, where M is the number of classical shadows. In other words, the number of experiments needed to achieve the statistical error ϵ is given by

$$M \geq \text{Var}(\hat{o})/\epsilon\delta^2. \quad (\text{A6})$$

Therefore, the sample complexity is directly related to the variance of single-shot estimation $\text{Var}(\hat{o})$. We can further bound the variance by

$$\begin{aligned} \text{Var}(\hat{o}) &= \mathbb{E}[\hat{o}^2] - \mathbb{E}[\hat{o}]^2 \leq \mathbb{E}[\hat{o}^2] \\ &= \mathbb{E}_{U \sim \mathcal{U}} \sum_{b \in \{0,1\}^n} \langle b|U\sigma U^\dagger|b\rangle \langle b|U\mathcal{M}^{-1}(O)U^\dagger|b\rangle^2 \\ &\leq \|O\|_{\text{shadow}}^2, \end{aligned} \quad (\text{A7})$$

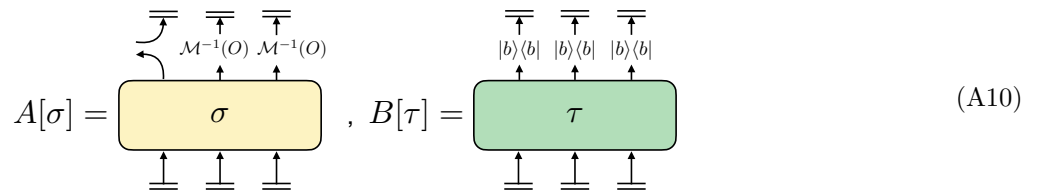
where the shadow norm of an observable is defined as

$$\begin{aligned} \|O\|_{\text{shadow}} &= \max_{\sigma: \text{state}} \left(\mathbb{E}_{U \sim \mathcal{U}} \sum_{b \in \{0,1\}^n} \langle b|U\sigma U^\dagger|b\rangle \langle b|U\mathcal{M}^{-1}(O)U^\dagger|b\rangle^2 \right)^{1/2} \\ &= \max_{\sigma: \text{state}} \left(\mathbb{E}_{U \sim \mathcal{U}} \sum_{b \in \{0,1\}^n} \text{Tr}(\sigma U^\dagger|b\rangle \langle b|U \langle b|U\mathcal{M}^{-1}(O)U^\dagger|b\rangle^2) \right)^{1/2} \\ &= \max_{\sigma: \text{state}} (\text{Tr} \sigma V_{\mathcal{U}}[O])^{1/2}, \end{aligned} \quad (\text{A8})$$

where we define a new operator $V_{\mathcal{U}}[O] = \mathbb{E}_{U \sim \mathcal{U}} \sum_{b \in \{0,1\}^n} U^\dagger|b\rangle \langle b|U \langle b|U\mathcal{M}^{-1}(O)U^\dagger|b\rangle^2$ that depends both on the unitary ensemble $\mathcal{U}$ and observable O . If the unitary ensemble $\mathcal{U}$ satisfies unitary 3-design, it can be simplified as

$$\begin{aligned} V_{\mathcal{U}}[O] &= \mathbb{E}_{U \sim \mathcal{U}} \sum_{b \in \{0,1\}^n} U^\dagger|b\rangle \langle b|U \langle b|U\mathcal{M}^{-1}(O)U^\dagger|b\rangle^2 \\ &= \sum_{b \in \{0,1\}^n} \sum_{\sigma, \tau \in S_3} \text{Wg}[\sigma\tau^{-1}g_0] A[\sigma] B[\tau], \end{aligned} \quad (\text{A9})$$

where σ, τ are permutations from permutation group S_3 , $\text{Wg}[g]$ is the Weingarten function of the permutation group element g , $g_0 = (2, 3)$ is a fixed permutation to match the tensor network connection, and $A[\sigma]$, $B[\tau]$ are defined as:



$$A[\sigma] = \text{[Diagram of } \sigma \text{ tensor network]}, \quad B[\tau] = \text{[Diagram of } \tau \text{ tensor network]} \quad (\text{A10})$$

In the following, we will mainly focus on the analysis of $V_{\mathcal{U}}[O]$ operator and $\|O\|_{\text{shadow}}^2$ with respect to different unitary ensembles $\mathcal{U}$.

3. Sample complexity of single logical qubit

Suppose a logical qubit is encoded using a $[n, k, l]$ stabilizer code, and $P = \prod_{g_i \in \mathcal{S}} \frac{1+g_i}{2}$ is the projection associated with the code. We choose the unitary ensemble used in shadow tomography to be $\mathcal{U} \sim \mathcal{C}\ell(2^n)$, which we dub the

global Clifford ensemble for logical qubits. The reconstruction channel of this ensemble is well-known,

$$\mathcal{M}^{-1}[A] = (2^n + 1)A - \text{Tr}(A)\mathbf{1}. \quad (\text{A11})$$

To get the corrected expectation of some observable O :

$$\langle O \rangle = \frac{\text{Tr}(P\rho PO)}{\text{Tr}(P\rho P)} = \frac{\text{Tr}(\rho PO)}{\text{Tr}(\rho P)}, \quad (\text{A12})$$

we need to estimate $\text{Tr}(\rho PO)$ and $\text{Tr}(\rho P)$ accurately. In other words, we need to use classical shadows to predict observable PO and P .

For projection operator P , $\mathcal{M}^{-1}(P) = (d+1)P - 2\mathbf{1}$, where $d = 2^n$ is the Hilbert space dimension associated with one logical qubit. We can calculate Eq. (A9) for the projection operator and get

$$V_{\mathcal{C}\ell(d)}[P] = \frac{2d-2}{d+2}(P + \mathbf{1}). \quad (\text{A13})$$

Therefore, the associated shadow norm is

$$\|P\|_{\text{shadow}}^2 = \max_{\sigma} \frac{2d-2}{d+2} \text{Tr}(\sigma P) + \frac{2d-2}{d+2} \text{Tr}(\sigma) \leq 4 \quad (\text{A14})$$

If O is a pure state fidelity operator, i.e. $O = \rho' = |\psi\rangle\langle\psi|$, we have $PO = O = \rho'$, $\text{Tr}(PO) = \text{Tr}(\rho') = 1$, and $\mathcal{M}^{-1}(PO) = (d+1)\rho' - \mathbf{1}$. With these, we can calculate

$$V_{\mathcal{C}\ell(d)}[\rho'] = \frac{d}{d+2}(2\rho' + \mathbf{1}). \quad (\text{A15})$$

Therefore, the shadow norm for fidelity operator is

$$\|\rho'\|_{\text{shadow}}^2 = \max_{\sigma} \frac{d}{d+2} (2\text{Tr}(\rho'\sigma) + \text{Tr}(\sigma)) \leq 3. \quad (\text{A16})$$

If O is a non-trivial Pauli operator for the logical qubit, then $\text{Tr}(PO) = 0$, $\mathcal{M}^{-1}(PO) = (d+1)PO$, and

$$V_{\mathcal{C}\ell(d)}[PO] = \frac{2d+2}{d+2}(PO + \mathbf{1}). \quad (\text{A17})$$