Logical shadow tomography: Efficient estimation of error-mitigated observables

Hong-Ye Hu,[1, 2, 3, 4, *] Ryan LaRose,[5, 3, *] Yi-Zhuang You,[1] Eleanor Rieffel,[3] and Zhihui Wang[3, 4, †]

[1]*Department of Physics, University of California San Diego, La Jolla, CA 92093, USA*
[2]*Department of Physics, Harvard University, 17 Oxford Street, Cambridge, MA 02138, USA*
[3]*Quantum Artificial Intelligence Laboratory (QuAIL),*
NASA Ames Research Center, Moffett Field, CA, 94035, USA
[4]*USRA Research Institute for Advanced Computer Science (RIACS), Mountain View, CA, 94043, USA*
[5]*Department of Computational Mathematics, Science, and Engineering,*
Michigan State University, East Lansing, MI, 48823, USA

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 a general nonlinear function of the noisy density matrix into the codespace. The subspace expansion and virtual distillation can be viewed as special cases of the new framework. We show our method is favorable in the quantum and classical resources overhead. Relative to subspace expansion which requires $O\left(2^N\right)$ samples to estimate a logical Pauli observable with $[[N, k]]$ error correction code, our technique requires only $O\left(4^k\right)$ samples. Relative to virtual distillation, our technique can compute powers of the density matrix without additional copies of quantum states or quantum memory. We present numerical evidence 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 [1–6]. 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) [7].

A relatively large number of QEM techniques have been proposed in recent literature [8–15]. Some authors have defined common frameworks which encapsulate one or more of these techniques [16, 17]. 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 are interested in error mitigation in the region between noisy-intermediate scale quantum (NISQ) and fault-tolerant quantum computation. Especially, we present a conceptually simple and practical QEM technique. In our setup, we use quantum error correction code to distribute logical qubits information with multiple physical qubits. Without active quantum error correction which requires extra ancilla qubits and parity check measurements, the only step which happens on a quantum computer is repeatedly sampling from an encoded state to perform shadow tomography [18, 19]. After this, one can use classical post-processing to project out errors as in subspace expansion [10] and calculate powers of the density matrix as in virtual distillation [12]. In addition to enabling both of these methods, our technique requires significantly fewer resources. Especially,

we show 1) the quantum gate overhead is independent of the number of logical qubits and we do not require multiple copies of the physical systems, 2) the sample complexity for estimating error mitigated Pauli observables only scales with the number of logical qubits instead of the total number of physical qubits, and 3) there exists an efficient classical algorithm that can post-process data with polynomial time. We show the new method is practical for relatively large systems with numerical evidence.

II. LOGICAL SHADOW TOMOGRAPHY

A. Motivation

One central task in quantum information processing is to estimate the expectation value of an observable O with respect to a pure state $\rho = |\psi\rangle\langle\psi|$, i.e. $\langle O \rangle = \langle\psi|O|\psi\rangle = \mathrm{Tr}(\rho O)$. If the state is prepared by a quantum device, it can be noisy, and we instead evaluate $\langle O \rangle_{\mathrm{noisy}} = \mathrm{Tr}(\rho_{\mathcal{E}} O)$ by direct measurement, where $\rho_{\mathcal{E}} = \mathcal{E}(\rho)$ denotes the noisy state that is corrupted from the target state ρ by an unknown noisy quantum channel $\mathcal{E}$. Given the corrupted state $\rho_{\mathcal{E}}$, the goal of quantum error mitigation is to estimate a quantity $\langle O \rangle_{\mathrm{QEM}}$ that is closer to the target value $\langle O \rangle$ compared to the direct measurement, such that

$$|\langle O \rangle_{\mathrm{QEM}} - \langle O \rangle| < |\langle O \rangle_{\mathrm{noisy}} - \langle O \rangle|. \qquad (1)$$

QEM can be performed on either physical devices or during data post-processing. In this work, we are interested in error mitigation in the region between noisy intermediate-scale quantum (NISQ) and fault-tolerant quantum computation, where the number of qubits is
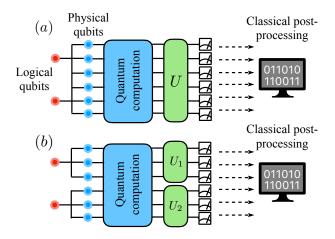

FIG. 1. Graphic illustration of logical shadow tomography.
(a) Red dots are logical qubits, and blue dots are physical
qubits. Logical information is distributed to physical qubits
by error correction code, then followed by noisy quantum com-
putation on physical qubits. To get estimation of error mit-
igated observables, we perform classical shadow tomography
on the noisy physical state. Particularly, we can apply random
Clifford gates denoted as green blocks from some unitary en-
semble $\mathcal{U}$, and take computational basis measurements. (b)
A special case using $[[n,1]]$ code for each logical qubit. In
shadow tomography, we apply random unitary from tensor
product of Clifford groups $\mathcal{Cl}(2^n)^{\otimes k}$. Additional gate depth
will not scale with number of logical qubits k, and sample
complexity for estimating error mitigated logical Pauli ob-
servables is the same as using global Clifford group $\mathcal{Cl}(2^{nk})$.

more than a few but cannot fulfill the requirement of
fault-tolerance. We ask the question whether we can use
those resources cleverly to achieve a more reliable quan-
tum computation. QEM methods related to this idea are
subspace expansion [10] and the virtual distillation [12].
The key contribution of this work is to propose the appli-
cation of classical shadow tomography [18, 19] to reduce
the quantum and classical resources needed to perform
these QEM schemes, and provide rigorous analysis on its
error mitigation capability.

The subspace expansion approach starts by encoding k
logical qubits with N physical qubits via an error correc-
tion code $[[N,k]]$. Consider using a stabilizer code defined
by a stabilizer group $\mathcal{S} = \langle S_1, S_2, \cdots, S_{N-k} \rangle$, the code
subspace is specified by the projection operator

$$\Pi = \prod_{j=1}^{N-k} \frac{\mathbb{1} + S_j}{2} = \frac{1}{2^{N-k}} \sum_{M \in \mathcal{S}} M, \quad (2)$$

where M denote group elements in $\mathcal{S}$ as products of the
stabilizers. Errors may occur to the physical state dur-
ing the quantum information processing, which generally
takes the physical state away from the code subspace.
Suppose the goal is to estimate the logical observable O
only, then even without active error correction, a simple
projection of the the corrupted physical state $\rho_\mathcal{E}$ back to
the code subspace can already mitigate the error for the

logical observable [10]

$$\langle O \rangle_{\text{QEM}} = \frac{\text{Tr}(\Pi \rho_\mathcal{E} \Pi^\dagger O)}{\text{Tr}(\Pi \rho_\mathcal{E} \Pi^\dagger)} = \frac{1}{2^{n-k}c} \sum_{M \in \mathcal{S}} \text{Tr}(\rho_\mathcal{E} MO), \quad (3)$$

where $c = \text{Tr}(\Pi \rho_\mathcal{E} \Pi^\dagger)$. This amounts to measuring $\langle MO \rangle$
on the corrupted physical state for all elements M in the
stabilizer group $\mathcal{S}$ (or for a majority of M sampled from
$\mathcal{S}$). This approach can quickly become exponentially ex-
pensive when $N - k$ becomes large.

Another approach for QEM goes under the name of vir-
tual distillation. Assuming the target state is a pure state
as the leading eigen state of $\rho_\mathcal{E}$. The sub-leading eigen
states of $\rho_\mathcal{E}$ (as errors orthogonal to the target state)
can be suppressed by powering the density matrix $\rho_\mathcal{E}^m$,
and we estimate $\langle O \rangle_{\text{QEM}} = \text{Tr}(\rho_\mathcal{E}^m O)/\text{Tr}(\rho_\mathcal{E}^m)$. Or more
generally, a polynomial function $f(\rho_\mathcal{E}) = c_0 \mathbb{1} + c_1 \rho_\mathcal{E} +
c_2 \rho_\mathcal{E}^2 + \cdots + c_m \rho_\mathcal{E}^m$ of $\rho_\mathcal{E}$ can be considered, with an opti-
mal choice of the coefficients $c_0, \cdots, c_m$ to best mitigate
the error, such that

$$\langle O \rangle_{\text{QEM}} = \frac{\text{Tr}(f(\rho_\mathcal{E})O)}{\text{Tr}(f(\rho_\mathcal{E}))}. \quad (4)$$

However, powering a density matrix on quantum devices
typically involves making multiple copies of the quantum
system, which can be challenging and expensive in quan-
tum resources.

The key observation of this work is that both Eq. (3)
and Eq. (4) (or their combination) can be efficiently eval-
uated in the classical post-processing phase after per-
forming the classical shadow tomography [19] on the
noisy physical state $\rho_\mathcal{E}$. The classical shadow tomog-
raphy uses randomized measurements to extract infor-
mation from an unknown quantum state, and predicts
physical properties about the state efficiently by post-
processing collected measurement outcomes on a classical
computer. When the measurement basis are chosen to be
Pauli or Clifford basis, the post-processing can be made
efficient. The code subspace projection and the powering
of density matrix can all be implemented efficiently in the
post-processing phase by classical computation, given the
Clifford nature of the classical shadows. In this way, we
can implement the existing error mitigation schemes with
significantly reduced quantum and classical resources.

B. Procedure

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

---

[1] Steps 2 and 3 can be aptly summarized by "perform shadow
tomography [19] on the logical state," whence *logical shadow
tomography*. We explain these steps in detail for completeness.

1. Given a k-qubit logical state, encode it into a N-qubit physical state by a $[[N, k]]$ stabilizer code.

2. Perform the quantum information processing on the physical state, the resulting physical state is denoted as $\rho_{\mathcal{E}}$. Due to the error accumulated in the processing, $\rho_{\mathcal{E}} = \mathcal{E}(\rho)$ may be corrupted from the ideal result $\rho = |\psi\rangle\langle\psi|$ by some noisy quantum channel $\mathcal{E}$. The goal is the mitigate the error for predicting logical observables based on the noisy physical state $\rho_{\mathcal{E}}$.

3. Perform classical shadow tomography on the physical state.

   3.1. Apply a randomly sampled unitary U from a unitary ensemble $\mathcal{U}$ to the physical qubits. Measure all physical qubits in the computational basis to obtain a bit-string $b \in \{0, 1\}^n$. Store U, b.

   3.2. Repeat step 3.1 for N times to obtain a data ensemble $\{(U_s, b_s)\}_{s=1}^N$ (s labels the samples in the ensemble).

4. Post-processing the data on a classical computer.

   4.1. Construct the classical shadow ensemble

   $$\Sigma(\rho_{\mathcal{E}}) = \{\hat{\rho}_s = \mathcal{M}^{-1}(U_s^\dagger |b_s\rangle\langle b_s|U_s)\}_{s=1}^N, \qquad (5)$$

   where $\mathcal{M}^{-1}$ is the classical shadow reconstruction map that depends on the unitary ensemble $\mathcal{U}$.

   4.2. Given any logical observable O (i.e. $[\Pi, O] = 0$) estimate the error-mitigated expectation value by

   $$\langle O \rangle_{\text{LST}} = \frac{\text{Tr}(\Pi f(\rho_{\mathcal{E}})\Pi^\dagger O)}{\text{Tr}(\Pi f(\rho_{\mathcal{E}})\Pi^\dagger)}, \qquad (6)$$

   where Π is the code subspace projection operator defined in Eq. (2), and $f(x) = \sum_{p=1}^m c_p x^p$ can be a generic polynomial function up to the power m. The proposed QEM estimator $\langle O \rangle_{\text{LST}}$ in Eq. (6) combines the subspace expansion Eq. (3) and the virtual distillation Eq. (4) approaches. In particular, the numerator $\text{Tr}(\Pi f(\rho_{\mathcal{E}})\Pi^\dagger O)$ is evaluated by

   $$\sum_{p=1}^m c_p \mathop{\mathbb{E}}_{\{\hat{\rho}_s\} \in \Sigma(\rho_{\mathcal{E}})^{\times p}} \text{Tr}\left(\Pi\Big(\prod_{s=1}^p \hat{\rho}_s\Big)\Pi^\dagger O\right), \qquad (7)$$

   and the denominator $\text{Tr}(\Pi f(\rho_{\mathcal{E}})\Pi^\dagger)$ is evaluated independently in a similar manner (by replacing O with $\mathbb{1}$).

The map $\mathcal{M}^{-1}$ depends on the unitary ensemble $\mathcal{U}$ for which there are several proposals, e.g. Pauli ensembles, random Clifford circuits, and chaotic dynamics [19–22].

In this work, we find the sample complexity for predicting logical Pauli observable is the same between using a full Clifford ensemble $\mathcal{Cl}(2^N)$ as shown in Fig. 1(a) and a tensor product of Clifford ensemble $\mathcal{Cl}(2^{N/k})^{\otimes k}$ as shown in Fig. 1(b). In the following, we will focus on the scheme where each qubit is encoded with a $[[n, 1]]$ error correction code and apply random unitaries from Clifford group $\mathcal{Cl}(2^n)$ at each logical qubit sector. And the total number of physical qubits is $N = nk$.

C.   Analysis

In the previous section, we have outlined the procedure of logical shadow tomography (LST). Here, we will analyze its performance from three perspectives: 1) error mitigation capacity, 2) quantum resources, and 3) classical resources. Particularly, in <u>error mitigation capacity</u> subsection, we show how error is suppressed with the code distance d of the error correction code and powers of density matrix. In <u>quantum resources</u> subsection, we show the gate overhead is similar to the original proposal of subspace expansion, except for a shallow depth Clifford circuit whose depth does not depend on the number of logical qubits. Compared to virtual distillation, our method only requires one copy of the physical system. In addition, we also show the sample complexity has an exponential reduction compared to the direct implementation of subspace expansion for estimating logical Pauli observables. In <u>classical resources</u> subsection, we outline the general classical algorithm for post-processing the data. Particularly, we show there exists fast algorithm for LST with $f(\rho_{\mathcal{E}}) = \rho_{\mathcal{E}}$ and its algorithm time complexity is $O(N^3)$. This allows our method scale to large system size.

1.   Error mitigation capability

• **Code space projection**. 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

$$\Pi E|\psi\rangle \propto (\mathbb{1} + S)E|\psi\rangle = E(|\psi\rangle - |\psi\rangle) = 0. \qquad (8)$$

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

$$\Pi|\psi\rangle = |\psi\rangle. \qquad (9)$$

Thus the projector Π discards results in which correctable errors have occurred [10]. The set of correctable errors is determined by the chosen code. Assume a simple noise model where each qubit is subjected to depolarizing noise with rate p. If a single error happened on one qubit. Then it can be projected out given the fact that single Pauli operator is not the stabilizer group. Those errors are non-logical errors.

The code space projection fails when more local error happens and they form a logical operator. The probability of having this failure is p^d, where d is the code distance of the error correction code. Mathematically, we can write down

$$\rho_{\mathcal{E}} = (1-p)^N \rho_0 + p\rho_1 + p^2\rho_2 + \dots, \quad (10)$$

where $\rho_0 = |\psi_0\rangle\langle\psi_0|$ is the ideal quantum state, and N is the system size. ρ_1 is the quantum state subjected to error happened to one qubit, i.e.

$$\rho_1 = (XI\cdots I)\rho_0(XI\cdots I) + (IX\cdots I)\rho_0(IX\cdots I) + \dots \quad (11)$$

Similarly, ρ_2 is the quantum state subjected to two error happened, i.e.

$$\rho_2 = (XIY\cdots I)\rho_0(XIY\cdots I) \\ + (IXX\cdots I)\rho_0(IXX\cdots I) + \dots \quad (12)$$

Therefore, for any logical observable O, we have

$$\frac{\mathrm{Tr}(\Pi\rho_{\mathcal{E}}\Pi O)}{\mathrm{Tr}(\Pi\rho_{\mathcal{E}}\Pi)} \\ = \langle\psi_0|O|\psi_0\rangle\left[1 + O\left(p^d\frac{\mathrm{Tr}(\rho_d O)}{\langle\psi_0|O|\psi_0\rangle}\right)\right]. \quad (13)$$

More details can be found in Appendix B.

• **Virtual distillation**. For completeness, we will review the virtual distillation theory here. With the spectral decomposition of the noisy density matrix, one can write

$$\rho_{\mathcal{E}} = p_0|\psi_0\rangle\langle\psi_0| + p_1|\psi_1\rangle\langle\psi_1| + \dots + p_n|\psi_n\rangle\langle\psi_n|, \quad (14)$$

where $p_0 > \dots > p_n$ and $\langle\psi_i|\psi_j\rangle = \delta_{ij}$. For shallow circuit, it is reasonable to assume $|\psi_0\rangle \equiv |\psi\rangle$ is the noiseless state.

In this case, for any observable O and positive integer m we have [12, 16]

$$\frac{\mathrm{Tr}(\rho_{\mathcal{E}}^m O)}{\mathrm{Tr}(\rho_{\mathcal{E}}^m)} = \frac{p_0^m\langle\psi_0|O|\psi_0\rangle + \sum_i p_i^m\langle\psi_i|O|\psi_i\rangle}{p_0^m + \sum_i p_i^m} \\ = \langle\psi|O|\psi\rangle\left[1 + O\left(\left(\frac{p_1}{p_0}\right)^m\frac{\langle\psi_1|O|\psi_1\rangle}{\langle\psi|O|\psi\rangle}\right)\right]. \quad (15)$$

Thus, computing the expectation value with the mth power of the state suppresses errors exponentially in m, a phenomenon which can also be interpreted as artificially cooling the system [23]. Here we allow for an arbitrary function f acting on the noisy state via its Taylor expansion $f(x) = \sum_{p=1}^m c_p x^p$. Including sums of powers up to m (instead of just the highest power m) was shown to improve results of numerical experiments in [16].

• **LST (Combined approach)**. LST has the error mitigation capability of code space projection and virtual distillation. When we combine code space projection with virtual distillation, we expect both code distance d

of the error correction code and mth power of the density matrix will suppress the error. Especially, when $m = 2$ in $\mathrm{Tr}(\Pi\rho_{\epsilon}^m\Pi O)$, i.e. one projects the squared density matrix to code space, the order of error suppression is improved from $\mathcal{O}(p^d)$ to $\mathcal{O}(p^{2d})$. In general, higher order of the power m will lead to stronger error mitigation effect. (See Appendix B for more details.)

2. Quantum resources

• **Gate overhead**. LST requires additional qubits and gates to encode the logical state, the exact number of which depends on the chosen code. Restricting to stabilizer codes, the logical state preparation only requires implementing Clifford gates, which are presumably easier to implement on NISQ devices compared to universal quantum computation gate set. This encoding overhead is the same as the subspace expansion method. To perform the classical shadow tomography, an element from a unitary ensemble $\mathcal{U}$ is appended to the quantum circuit before measuring all qubits in the computational basis. If $\mathcal{U}$ is chosen to be a global Clifford ensemble, the added circuit depth is $\mathcal{O}(N)$ with local random unitary gates [24]. The gate overhead associated with this can be significant. If $\mathcal{U}$ is chosen to be a tensor-product Pauli ensemble, the added circuit depth is $\mathcal{O}(1)$ [19]. The gate overhead is more affordable. However, the Pauli measurement will increase the sample complexity exponentially for non-local observables. Facing this dilemma between Clifford v.s. Pauli measurement, a recent work [21] by part of the authors has developed an efficient classical shadow tomography approach for finite-depth local Clifford circuits, which can smoothly interpolate between the global Clifford and the local Pauli limit. Using shallow (finite-depth) Clifford circuits for shadow tomography, the gate overhead can be significantly reduced with only a mild increase of sample complexity. [21] This approach can be combined with our error mitigation technique seamlessly to achieve an optimal balance between gate overhead and sample complexity. For this work, we will use a tensor-product Clifford group $\mathcal{Cl}(2^n)^{\otimes k}$ as shown in Fig. 1 (b). The added circuit depth is $\mathcal{O}(n)$, where n is the physical qubits in $[[n,1]]$ code for one logical qubit. It is important to notice that the depth of additional circuit is independent of the number of logical qubit k.

• **Sample complexity**. We now consider number of measurements needed for classical post-processing in LST. If one uses a $[[n,1]]$ code for each logical qubit and a tensor-product Clifford group $\mathcal{Cl}(2^n)^{\otimes k}$, where k is number of logical qubits, we have the following theorem which dictates the sample complexity of LST.

Theorem 1 (Details in Appendix D) *Let ρ be a k logical qubits quantum state, where each logical qubit is encoded with a $[[n,1]]$ code, and Π be the associated projection operator to the code subspace. Then one needs $\mathcal{O}(\log(M)4^k/\epsilon\delta^2)$ samples to produce an estimation $\tilde{O}_i$*

of $\mathrm{Tr}(\Pi\rho\Pi O_i)$ with $\{i = 1, \ldots, M\}$ of logical Pauli observables O_i such that

$$Pr\left(\left|\tilde{O}_i - \mathrm{Tr}(\Pi\rho\Pi O_i)\right| \geq \epsilon\right) \leq \delta. \qquad (16)$$

The result (16) applies to both the numerator and denominator of the LST estimate Eq. (6) separately. We emphasize that the number of samples does not depend on the total number of physical qubits nk, but only depends on the number of logical qubits k, and scales as $\mathcal{O}(4^k)$. Compared to direct implementation of subspace expansion whose sample complexity is $\mathcal{O}(2^{nk})$, our method dramatically reduces the sample complexity.

In Ref. [10], the authors also mentioned stochastic sampling of the stabilizer group elements to reduce the sample complexity at the price of losing the error mitigation capacity. Our approach bypasses the need to sample the all elements in the stabilizer group, as we can implement the projection operator directly and efficiently by data post-processing. Our advantage will be even more apparent if one uses larger n error correction code $[[n, 1]]$, which provides larger code distance and better error mitigation capacity.

3. Classical resources

In this section, we show that there exists efficient algorithm for classical post-processing. Especially, for LST with $f(\rho_{\mathcal{E}}) = \rho_{\mathcal{E}}$ (no virtual distillation), the classical post-processing can be performed with polynomial classical memory and time.

After sampling, we have the classical shadow Eq. (5) consisting of N stabilizer states $U^\dagger|b\rangle\langle b|U$, the observable O, and the projector Π. We need to estimate the numerator and denominator of Eq. (6), both of which can be written

$$\mathrm{Tr}[f(\rho_{\mathcal{E}})\Gamma] \qquad (17)$$

where $\Gamma = \Pi O$ for the numerator and $\Gamma = \Pi$ for the denominator. Let m be the highest degree of the Taylor expansion of f. As in [19], we can evaluate the expectation of this term via

$$\mathbb{E}\,\mathrm{Tr}[\hat{\rho}_{i_1}\cdots\hat{\rho}_{i_m}\Gamma], \qquad (18)$$

where $\{\hat{\rho}_{i_1}\ldots\hat{\rho}_{i_m}$ are independent samples. A single classical shadow $\hat{\rho}_i$ requires $\mathcal{O}(N^2)$ classical memory to store (where $N = nk$ is the total number of qubits), so the argument inside the trace in (18) requires $\mathcal{O}(N^2 m)$ classical memory [25]. In Appendix A 1, we showed that the evaluation of (18) boils down to evaluating the general form

$$\mathrm{Tr}\left(\prod_{j=1}^{l}(a_j\mathbb{1} + b_j M_j)\right), \qquad (19)$$

where M_j are Pauli operators. This can be solved by finding the null space $\mathcal{N}_A$ of a binary matrix representation of Pauli operators $\{M_j\}$. By simple counting argument (see Appendix A 1 for more details), we see the evaluation of Eq. (18) has time complexity upper bounded by $\mathcal{O}\left(mN^2(mN + N - k + 1) + m|\mathcal{N}_A|\right)$, when $m \geq 2$, where k is number of logical qubits, $N = nk$ is the total number of physical qubits, m is power of density matrix, and $|\mathcal{N}_A|$ is the volume of binary null space $\mathcal{N}_A$ determined by a set of classical shadows $\{\hat{\rho}_{i_1}, \ldots, \hat{\rho}_{i_m}\}$. When m is large, the null space $\mathcal{N}_A$ can be very large. In practice, the evaluation of (18) can be slow given the time complexity is proportional to $\mathcal{O}(m|\mathcal{N}_A|)$ for $m \geq 2$.

We also find an improved classical algorithm for $m = 1$ with time complexity $\mathcal{O}(N^3)$. When $m = 1$, we would like to evaluate $\mathrm{Tr}(\Pi\rho\Pi O)$, where ρ is proportional to a stabilizer state, which can be represented using stabilizer tableau. The intuition behind the efficient algorithm is that a stabilizer state can be efficiently projected by another stabilizer group projector by updating its stabilizer tableau. We leave the detail of the algorithm to Appendix A 2.

In conclusion, at least for the $m = 1$ (no virtual distillation) case, the post-processing complexity is polynomial $\mathcal{O}(N^3)$ in the total qubit number N.

III. NUMERICAL RESULTS

In Sec. II C, we discussed the error mitigation capabilities, quantum resources, and classical resources of LST. One should notice that the discussion of sample complexity is mainly focused on estimation of $\mathrm{Tr}(\Pi\rho\Pi^\dagger O)$. While in practice, we would like to estimate the ratio of $\mathrm{Tr}(\Pi\rho\Pi^\dagger O)/\mathrm{Tr}(\Pi\rho\Pi^\dagger)$. The sample complexity of the ratio does not have a closed form in general. This problem is not unique to our approach. It also exists for the subspace expansion [10] and virtual distillation [12].

Nevertheless, in the following, we will use numerical experiments to investigate the sample complexity. We demonstrate the performance of LST through numerical simulation of large systems. In particular, we find LST outperforms the direct implementation of the subspace expansion and the sample complexity scaling is very close to our theoretical prediction in small noise region.

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

We first consider a simple example with one logical qubit encoded in five physical qubits with the $[[5, 1, 3]]$ stabilizer code. Each physical qubit is subjected to depolarizing noise with error rate p. The same model is considered in the subspace expansion literature [10], which shows pseudo-threshold $p = 0.5$. Here, we want to compare the practical performance of logical shadow tomography and direct measurement by subspace expansion.
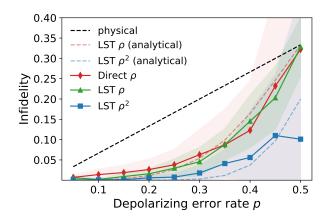
FIG. 2. LST with the $[[5, 1, 3]]$ code. Here, $|\psi\rangle$ is taken to be the logical $|\bar{0}\rangle$ and we estimate infidelity $1 - F$ with samples. The dashed black line shows the physical infidelity, i.e., the noisy expectation value of single qubit without any encoding. The green and blue dashed line are analytical performance of logical shadow tomography with $f(\rho_\mathcal{E}) = \rho_\mathcal{E}$ and $f(\rho_\mathcal{E}^2) = \rho_\mathcal{E}^2$ respectively. The red dots and red shaded area indicates the mean value and standard deviation of error mitigation with $f(\rho_\mathcal{E}) = \rho_\mathcal{E}$ by direct implementation of subspace expansion with 3000 measurements. The green line and green shaded area indicate the mean value and standard deviation with $f(\rho_\mathcal{E}) = \rho_\mathcal{E}$ and 3000 measurements by LST. And the performance of LST with $f(\rho_\mathcal{E}) = \rho_\mathcal{E}^2$ is indicated by blue line and blue shaded area.

We evaluate Eq. (6) with $f(\rho_\mathcal{E}) = \rho_\mathcal{E}$ and $f(\rho_\mathcal{E}) = \rho_\mathcal{E}^2$. The results are shown in Fig. 2. Here, the dashed black line shows the infidelity without any error mitigation (the "physical" curve), and the dashed colored lines 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 $f(\rho_\mathcal{E}) = \rho_\mathcal{E}^2$ outperforms LST with $f(\rho_\mathcal{E}) = \rho_\mathcal{E}$ showing that the combination of codespace projection and virtually distillation outperforms only projecting into the codespace. This phenomenon agrees with the behavior of the error mitigation capability p^{md} with $m = 1, 2$ argued in II C 1 (See Appendix B for proof details). In addition to expected performance, we also care about sample efficiency, since one major contribution of our work is showing the exponential reduction in sample complexity with LST.

To show this practical advantage, we collect 3000 measurements results and use the data to estimate error mitigated value. The colored lines/points shows the mean values of the estimation and the colored regions shows the standard deviation to the mean value. For $f(\rho_\mathcal{E}) = \rho_\mathcal{E}$, LST is the same as the subspace expansion. If we implement the direct subspace expansion by measuring every Pauli observables, each Pauli observable is measured around 100 times. The red points/line and red shaded region shows the result of direct implementation of subspace expansion. Given each Pauli observable is not measured many times, small error in the denominator can cause large error of the ratio. We see the standard devi-

ation is very huge. The mean estimation and standard deviation of LST is shown as green points/line and green shaded region. In contrast to direct implementation of subspace expansion, LST has a much smaller fluctuation with the same amount of measurements. This shows the practical advantage of our method. In addition, the blue points/line and blue shaded region shows the result of LST with $f(\rho_\mathcal{E}) = \rho_\mathcal{E}^2$. With the same amount of data, we see it will suppress the noise even more.

B. Convergence vs. code size

As we have pointed out in Sec. II C 2, the sample complexity for estimating $\text{Tr}(\Pi\rho\Pi^\dagger O)$ for logical Pauli observables O only scales with number of logical qubits k as $O(4^k)$ but does not scale with the number of encoding qubits n of the $[[n, 1]]$ code for each logical qubit. In practice, we will estimate error mitigated values as a ratio, i.e. $\text{Tr}(\Pi\rho\Pi^\dagger O)/\text{Tr}(\Pi\rho\Pi^\dagger)$. Since there is no closed form for the statistical fluctuation of the above ratio. We will investigate the sample complexity of it via numerical simulation. Interestingly, we find the sample complexity of the ratio agrees well with our theoretical analysis in small noise region.

We now consider $[[n, 1]]$ codes for one logical qubit and vary the number of physical qubits ranging from $n = 10$ to $n = 60$. For quantum noise, each physical qubit is subjected to 1% depolarizing noise in all the following experiments. The LST estimated fidelity (using $f(\rho_\epsilon) = \rho_\epsilon$ for all code sizes) vs. number of samples is shown in Fig. 3 (a). 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} samples, a number infeasible to implement in any practical experiment. In addition, we estimate the standard deviation of each predicted point by LST using the bootstrap method, and the result is shown in Fig. 3 (b). We see the standard deviation does not show strong dependence of number of encoding qubits n, and it indicates the sample complexity does not increase much if one increase n.

In addition, we also study the sample complexity scaling with number of logical qubits k. Particularly, we use $[[5, 1, 3]]$ code to encode each logical qubit and prepare a multi-logical qubits GHZ state, $(|\bar{0}\ldots\bar{0}\rangle + |\bar{1}\ldots\bar{1}\rangle)/\sqrt{2}$. In Fig. 3 (c), the blue dots/line shows the estimated mean value for logical operator $\bar{X}_1 \otimes \cdots \otimes \bar{X}_k$, and the blue shaded area indicates the standard deviation. Especially, in the inset, we see the standard deviation increases exponentially with number of logical qubits k. This result also agrees well with our sample complexity analysis, even though the analysis focuses on estimating $\text{Tr}(\Pi\rho\Pi^\dagger O)$.

Through the large scale numerical simulation, we find LST indeed outperforms previous methods in terms of sample complexity. Interestingly, we also find the sample complexity agrees well with the theoretical analysis in
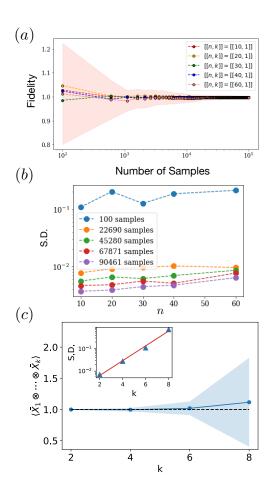
FIG. 3. Scaling study of LST with $f(\rho_\epsilon) = \rho_\epsilon$. In all figures, each physical qubit is subjected to 1% depolarizing noise. (a) LST estimated fidelity vs. number of samples from 10^2 - 10^5 with various $[[n, 1]]$ code sizes. The noiseless fidelity value of 1.0 is shown with the 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 [26].(b) Standard deviation vs. number of physical qubits n. The standard deviation of estimation doesn't scale with number of encoding physical qubits. (c) Mean value and standard deviation scaling vs. number of logical qubits k. Each logical qubit is encoded with $[[5, 1, 3]]$ code, and the state is prepared as logical GHZ state $|\bar{0} \ldots \bar{0}\rangle + |\bar{1} \ldots \bar{1}\rangle$. We see standard deviation scales exponentially with number of logical qubits k as predicted.

small noise region. This indicates the sample complexity of LST does not scale much with the number of physical encoding qubits n with $[[n, 1]]$ error correction code and only scales with number of logical qubits k.

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 Clifford 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. In the analysis, we show if the error correction code has code distance d and $f(\rho_\mathcal{E}) = \rho_\mathcal{E}$ for LST, then error will be suppressed to $\mathcal{O}(p^d)$, assuming independent depolarizing noise with rate p on each physical qubit. And we also show higher power of density matrix will further improve the performance. For sample complexity, we show it scales only with number of logical qubits k but not the number of encoding physical qubits n. This result is also supported with large scale numerics. In addition, we provide efficient classical algorithms to post-process the classical shadow data and remark that this post-processing 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 [10, 12] 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 [10, 16]. We have also implemented virtual distillation [12] 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 $f(\rho_\mathcal{E})$ of the noisy state $\rho_\mathcal{E}$; the only difference is in classical post-processing (and number of shadows needed). Note that virtual distillation without subspace expansion can be implemented in our protocol by using a trivial code (i.e., not encoding a logical state). Further, beyond making both of these techniques significantly more practical to implement, our procedure enables them to be composed with one another, and we have shown numerically the composition of both techniques results in further reduction of errors. Additional error mitigation techniques which act on the noisy state, e.g., those in [16], may also be implementable with our framework.

In our analysis, we assumed the Clifford circuit in the classical shadow part is noise-free. If there are noise in the Clifford circuit part, it can be mitigated if the noise is independent of the Clifford gates, as in [27–29], where similar idea was used for randomized benchmarking [30, 31].

Shadow tomography since proposed in [19] has found a number of applications in quantum information processing, including the recently proposed process tomography [32], and avoiding barren plateau in variational quantum circuits [33]. This work constitutes an application in the error mitigation realm. We are optimistic our procedure will be effective on current and near-term quantum computers for a variety of experiments on relatively large systems.

In a recent literature [15], the authors discuss an error

mitigation method to implement low noise logical gates for error correction codes. In comparison, our method focuses on the error mitigation of the final quantum states of the logical qubits.

While we were finishing the writing of the paper we learned of related work "Shadow Distillation: Quantum Error Mitigation with Classical Shadows for Near-Term Quantum Processors" by Alireza Seif, Ze-Pei Cian, Sisi Zhou, Senrui Chen, and Liang Jiang [34].

[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 Mruczkiewicz, 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 Marusczyk, Sam McArdle, Jarrod R. McClean, Trevor McCourt, Matt McEwen, Anthony Megrant, Carlos Mejuto-Zaera, Xiao Mi, Masoud Mohseni, Wojciech Mruczkiewicz, 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 Mruczkiewicz, 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] Sepehr Ebadi, Tout T. Wang, Harry Levine, Alexander Keesling, Giulia Semeghini, Ahmed Omran, Dolev Bluvstein, Rhine Samajdar, Hannes Pichler, Wen Wei Ho, Soonwon Choi, Subir Sachdev, Markus Greiner, Vladan Vuletić, and Mikhail D. Lukin. Quantum phases of matter on a 256-atom programmable quantum simulator. Nature, 595(7866):227–232, 2021.

[6] G. Semeghini, H. Levine, A. Keesling, S. Ebadi, T. T. Wang, D. Bluvstein, R. Verresen, H. Pichler, M. Kalinowski, R. Samajdar, A. Omran, S. Sachdev, A. Vishwanath, M. Greiner, V. Vuletić, and M. D. Lukin. Probing topological spin liquids on a programmable quantum simulator. Science, 374(6572):1242–1247, 2021.

[7] 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.

[8] Ying Li and Simon C. Benjamin. Efficient Variational Quantum Simulator Incorporating Active Error Minimization. Physical Review X, 7(2):021050, June 2017.

[9] 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.

[10] 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.

[11] 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.

[12] 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.

[13] Christophe Piveteau, David Sutter, Sergey Bravyi, Jay M. Gambetta, and Kristan Temme. Error mitigation for universal gates on encoded qubits. Physical Review Letters, 127(20), Nov 2021.

[14] Angus Lowe, Max Hunter Gordon, Piotr Czarnik, Andrew Arrasmith, Patrick J. Coles, and Lukasz Cincio. Unified approach to data-driven quantum error mitigation. Phys. Rev. Research, 3:033098, Jul 2021.

[15] Andrew Thomas Arrasmith, Piotr Jan Czarnik, Patrick Joseph Coles, and Lukasz Cincio. Error mitigation with clifford quantum-circuit data. Quantum, 5, 11 2021.

[16] 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.

[17] 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.

[18] Scott Aaronson. Shadow Tomography of Quantum States. arXiv:1711.01053 [quant-ph], November 2018.

arXiv: 1711.01053.

[19] 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.

[20] Hong-Ye Hu and Yi-Zhuang You. Hamiltonian-driven shadow tomography of quantum states. Phys. Rev. Research, 4:013054, Jan 2022.

[21] Hong-Ye Hu, Soonwon Choi, and Yi-Zhuang You. Classical Shadow Tomography with Locally Scrambled Quantum Dynamics. arXiv:2107.04817 [cond-mat, physics:quant-ph], July 2021. arXiv: 2107.04817.

[22] M Ohliger, V Nesme, and J Eisert. Efficient and feasible state tomography of quantum many-body systems. New Journal of Physics, 15(1):015024, jan 2013.

[23] 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.

[24] Fernando G. S. L. Brandao, Aram W. Harrow, and Michal Horodecki. Local random quantum circuits are approximate polynomial-designs. Communications in Mathematical Physics, 346(2):397–434, September 2016. arXiv: 1208.0692.

[25] Scott Aaronson and Daniel Gottesman. Improved Simulation of Stabilizer Circuits. Physical Review A, 70(5):052328, November 2004. arXiv: quant-ph/0406196.

[26] Markus Grassl. Bounds on the minimum distance of linear codes and quantum codes.

[27] Senrui Chen, Wenjun Yu, Pei Zeng, and Steven T. Flammia. Robust shadow estimation. PRX Quantum, 2(3), Sep 2021.

[28] Dax Enshan Koh and Sabee Grewal. Classical Shadows with Noise. arXiv e-prints, page arXiv:2011.11580, November 2020.

[29] Kaifeng Bu, Dax Enshan Koh, Roy J. Garcia, and Arthur Jaffe. Classical shadows with Pauli-invariant unitary ensembles. arXiv e-prints, page arXiv:2202.03272, February 2022.

[30] Easwar Magesan, J. M. Gambetta, and Joseph Emerson. Scalable and robust randomized benchmarking of quantum processes. Phys. Rev. Lett., 106:180504, May 2011.

[31] Jahan Claes, Eleanor Rieffel, and Zhihui Wang. Character randomized benchmarking for non-multiplicity-free groups with applications to subspace, leakage, and matchgate randomized benchmarking. PRX Quantum, 2:010351, Mar 2021.

[32] 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.

[33] Stefan H. Sack, Raimel A. Medina, Alexios A. Michailidis, Richard Kueng, and Maksym Serbyn. Avoiding

barren plateaus using classical shadows. arXiv e-prints, page arXiv:2201.08194, January 2022.

[34] Alireza Seif, Ze-Pei Cian, Sisi Zhou, Senrui Chen, and Liang Jiang. Shadow distillation: Quantum error miti-

gation with classical shadows for near-term quantum processors. (To appear).

Appendix A: Stabilizer algorithms

1. Evaluating the trace in Eq. (7)

In this appendix, we explain an efficient approach to evaluate the trace in Eq. (7). For random Clifford ensemble, the reconstruction map reads $\hat{\rho} = \mathcal{M}^{-1}(\hat{\sigma}) = (2^n + 1)\hat{\sigma} - \mathbb{1}$, such that

$$\mathbb{E}_{\hat{\rho}} \operatorname{Tr}\left(\Pi\Big(\prod_{s=1}^{m}\hat{\rho}_s\Big)\Pi^\dagger O\right) = \sum_{q=0}^{m}\binom{m}{q}(2^n+1)^q(-1)^{m-q}\mathbb{E}_{\hat{\sigma}}\operatorname{Tr}\left(\Pi\Big(\prod_{s=1}^{q}\hat{\sigma}_s\Big)\Pi^\dagger O\right). \tag{A1}$$

Notice that the projection operator $\Pi = \prod_{j=1}^{n-k}(\mathbb{1}+S_j)/2$ and the snapshot state $\hat{\sigma} = U^\dagger|b\rangle\langle b|U = \prod_{i=1}^{n}(\mathbb{1}+b_i U^\dagger Z_i U)/2$ both take the from of stabilizer states. So the problem boils down to evaluating the trace of the following general form

$$\operatorname{Tr}\left(\prod_{j=1}^{l}(a_j\mathbb{1} + b_j M_j)\right), \tag{A2}$$

where M_j are Pauli operators and a_j, b_j are real coefficients. As we expand the product, the only terms that survive the trace are those terms with the Pauli operators multiplied to the identity operator. To find these combination of Pauli operators, we can first encode every Pauli operator M_j as a binary vector following

$$M_j = \mathrm{i}^{\sum_{i=1}^{n}\xi_{ij}\zeta_{ij}}\prod_{i=1}^{n}X_i^{\xi_{ij}}\prod_{i=1}^{n}Z_i^{\zeta_{ij}} \rightarrow \begin{pmatrix}\vdots\\\xi_{ij}\\\zeta_{ij}\\\vdots\end{pmatrix}, \tag{A3}$$

where $\xi_{ij}, \zeta_{ij} \in \{0,1\}$ are binary variables. Arranging all the binary vector representations of M_j as column vectors, together they form a $2n \times l$ matrix, denoted as A. Each combination of Pauli operators M_j that multiply to identity corresponds to a binary null vector solution x of the binary matrix A, as $Ax = 0$ (modulo 2). The null vectors form the null space of A, denoted as $\mathcal{N}_A$. The null space of binary matrix A can be found using Gaussian elimination method, and its time complexity is $\mathcal{O}(nl \times \min(2n, l))$. For $x \in \mathcal{N}_A$,

$$\prod_{j=1}^{l}(M_j)^{x_j} = z(x)\mathbb{1}, \tag{A4}$$

which defines the phase factor $z(x)$ given x. Then the trace in Eq. (A2) is given by

$$\operatorname{Tr}\left(\prod_{j=1}^{l}(a_j\mathbb{1} + b_j M_j)\right) = 2^n\sum_{x\in\mathcal{N}_A}z(x)\prod_{j=1}^{l}a_j^{1-x_j}b_j^{x_j}. \tag{A5}$$

Therefore the time complexity of evaluating the general trace Eq. (A5) is $\mathcal{O}(nl \times \min(2n, l) + |\mathcal{N}_A|)$, where $|\mathcal{N}_A|$ is the volume of the null space $\mathcal{N}_A$ that is determined by the set of Pauli operators $\{M_i\}$. Applying this result for Eq. (A1), we get the time complexity for evaluating $\operatorname{Tr}\left(\Pi\left(\prod_{s=1}^{m}\hat{\rho}_s\right)\Pi^\dagger O\right)$ is upper bounded by $\mathcal{O}(mnl \times \min(2n, l) + m|\mathcal{N}_A|)$, with $l = mn + n - k + 1$. For large m, the volume of null space $|\mathcal{N}_A|$ can be troublesome, and scale exponentially with m. But luckily for $m = 1$, there exists more efficient polynomial time algorithm, which is illustrated in Appendix A 2.

2. Efficient projection of a stabilizer state

As shown in Eq. (A3), any Pauli string operator can be represented as a one-hot binary vector x and z, with $x_i, z_i = 0, 1$ for $i = 1, \ldots, N$, where N is the total number of qubits,

$$\sigma_{(x,z)} = i^{x \cdot z} \prod_{i=1}^{N} X_i^{x_i} Z_i^{y_i}, \tag{A6}$$

where X_i, Z_i are Pauli operators, and x_i, z_i are binary values. The multiplication of two Pauli operators can be represented as

$$\sigma_{(x,z)} \sigma_{(x',z')} = i^{p(x,z;x',z')} \sigma_{(x+x',z+z')\%2}, \tag{A7}$$

where the phase factor is

$$p(x, z; x', z') = \sum_{i=1}^{N} \left(z_i x_i' - x_i z_i' + 2(z_i + z_i') \left\lfloor \frac{x_i + x_i'}{2} \right\rfloor + 2(x_i + x_i') \left\lfloor \frac{z_i + z_i'}{2} \right\rfloor \right) \quad \text{mod } 4. \tag{A8}$$

Any two Pauli strings either commute or anti-commute,

$$\sigma_{(x,z)} \sigma_{(x',z')} = (-)^{c(x,z;x',z')} \sigma_{(x',z')} \sigma_{(x,z)}, \tag{A9}$$

where the anticommutation indicator c has a simpler form

$$c(x, z; x', z') = \frac{p(x, z; x', z') - p(x', z'; x, z)}{2} = \sum_{i=1}^{N} (z_i x_i' - x_i z_i') \quad \text{mod } 2. \tag{A10}$$

Therefore, the complexity of calculating anticommutation indicator is $\mathcal{O}(N)$. The binary vectors x and z can be interweaved into a $2N$-component vector $g = (x_0, z_0, x_1, z_1, \cdots)$, which forms the binary representation of a Pauli operator σ_g.
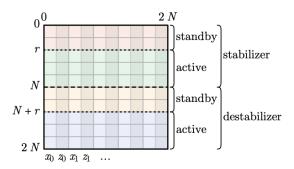

FIG. 4. Data structure of a stabilizer state. Each Pauli string is represented as a binary vector. First N rows store the stabilizers of the state, and second N rows store the destabilizers of the state.

In Fig. 4, each row is a binary representation of a Pauli string. For a Hilbert space with dimension $\mathbb{C}^{2^N}$, we can find at most N stabilizers $\{S_i, i = 1 \ldots N\}$ and N destabilizers $\{D_i, i = 1 \ldots N\}$, with $[S_i, S_j] = 0$, $[D_i, D_j] = 0$ and $\{S_i, D_j\} = \delta_{i,j}$. Fig. 4 is called the stabilizer tableau of stabilizer state. We use r in Fig. 4 indicates the $\log -2$ rank of the density matrix. For pure stabilizer state, $r = 0$. And if $r > 0$, then stabilizer tableau represents a mixed state and only partial Hilbert space is stabilized.

A stabilizer projector $\Pi = \prod_{k=1}^{l} \frac{\mathbb{1} \pm G_k}{2}$ can also be represented as a l-row tableau (only stabilizer). Now we are going to present an efficient algorithm to calculate $\text{Tr}(\Pi \rho \Pi)$, and update the stabilizer tableau of a full rank stabilizer state ρ according to projection Π.

• **Outline of the algorithm:** First, we set trace = 1. Then we scan over every observable G_k in the generator of the operator. For each G_k, we continue to scan over all operators in the stabilizer tableau. If the observable G_k anticommute with:

1. At least one active stabilizer (the first of them being S_p) $\rightarrow G_k$ is an <u>error</u> operator that take the state out of the code subspace $\rightarrow$ stabilizer tableau need to be updated according to $\overline{\mathbb{1} \pm G_k}$. And trace will be multiplied by $1/2$.

2. Otherwise, G_k is in the stabilizer group generated by $\{S_k\}$. If the phase factor is compatible, then stabilizer state is eigenstate of $(\mathbb{1} \pm G_k)/2$ with eigenvalue 1, i.e. $(\mathbb{1} \pm G_k)|\psi\rangle/2 = |\psi\rangle$; if the phase factor is incompatible, it means $(\mathbb{1} \pm G_k)|\psi\rangle/2 = 0$.

We see this algorithm has a double loop of $\mathcal{O}(N)$ items, and each anticommutation check takes $\mathcal{O}(N)$ time. Therefore, the time complexity of this algorithm is $\mathcal{O}(N^3)$, where N is the total number of qubits in the system.

Appendix B: Error mitigation capability

In Eq. (10), we show the density matrix subjected to depolarizing noise is naturally the spectral decomposition form. Here, we provide detailed proof of it. Let $|\psi_0\rangle$ be the ideal quantum state encoded with $[[n, k, d]]$ error correction code. And let $\{|\bar{i}\rangle, i = 1, \ldots, 2^k\}$ be the orthonormal basis for the logical space. In general, $|\psi_0\rangle = \sum_{i=1}^{2^k} c_i |\bar{i}\rangle$. We assume the simple depolarizing error for each physical qubit. Then

$$\rho_\epsilon = (1 - p)^N |\psi_0\rangle\langle\psi_0| + p\rho_1 + p^2\rho_2 + \cdots, \tag{B1}$$

where ρ_i is the density matrix with i local error happened, and N is the system size. For example,

$$\begin{aligned}\rho_1 =& (XII\cdots I)|\psi_0\rangle\langle\psi_0|(XII\cdots I) + (YII\cdots I)|\psi_0\rangle\langle\psi_0|(YII\cdots I) \\ &+ (ZII\cdots I)|\psi_0\rangle\langle\psi_0|(ZII\cdots I) + (IXI\cdots I)|\psi_0\rangle\langle\psi_0|(IXI\cdots I) \\ &+ \cdots + (II\cdots IZ)|\psi_0\rangle\langle\psi_0|(II\cdots IZ),\end{aligned} \tag{B2}$$

and

$$\begin{aligned}\rho_2 =& (XXI\cdots I)|\psi_0\rangle\langle\psi_0|(XXI\cdots I) + (XYI\cdots I)|\psi_0\rangle\langle\psi_0|(XYI\cdots I) \\ &+ \cdots + (I\cdots ZZ)|\psi_0\rangle\langle\psi_0|(I\cdots ZZ).\end{aligned} \tag{B3}$$

We define the support of a Pauli string as number of Pauli operators that is not the identity operator. Let P_l be a Pauli string operator with non-trivial support l. For example, the support of Pauli string $XIZYI$ is three. Then any term in ρ_l can be written as $P_l|\psi_0\rangle\langle\psi_0|P_l$ with some P_l. It is easy to check $\Pi P_l|\psi_0\rangle\langle\psi_0|P_l\Pi = 0$ for any $l < d$. By definition of the code distance d, any P_l with $l < d$ is not in the stabilizer group. Therefore, it must anti-commute with some stabilizer generator S, such that $\{S, P_l\} = 0$. We write $\Pi = \Pi'(\mathbb{1} + S)$, where Π' includes all other stabilizer generators. Then

$$\Pi P_l|\psi_0\rangle = \Pi'(\mathbb{1} + S)P_l|\psi_0\rangle = \Pi' P_l(\mathbb{1} - S)|\psi_0\rangle = \Pi' P_l(|\psi_0\rangle - |\psi_0\rangle) = 0, \tag{B4}$$

and

$$\Pi P_l|\psi_0\rangle\langle\psi_0|P_l\Pi = 0 \quad (l < d). \tag{B5}$$

Therefore, we conclude $\Pi\rho_l\Pi = 0$ for any $l < d$, and

$$\begin{aligned}&\frac{\text{Tr}(\Pi\rho_\mathcal{E}\Pi O)}{\text{Tr}(\Pi\rho_\mathcal{E}\Pi)} \\ &= \langle\psi_0|O|\psi_0\rangle\left[1 + O\left(p^d \frac{\text{Tr}(\rho_d O)}{\langle\psi_0|O|\psi_0\rangle}\right)\right].\end{aligned} \tag{B6}$$

Now we want to prove the leading order correction of $\Pi\rho_\epsilon^2\Pi$ is of order $\mathcal{O}(p^{2d})$ with contradiction. Suppose the leading order correction is of order $\mathcal{O}(p^s)$ with $s < 2d$, then there exist P_l and P_r with $l + r = s < 2d$ such that

$$\Pi(P_l|\psi_0\rangle\langle\psi_0|P_l)(P_r|\psi_0\rangle\langle\psi_0|P_r)\Pi = \langle\psi_0|P_l P_r|\psi_0\rangle(\Pi P_l|\psi_0\rangle))(\langle\psi_0|P_r\Pi) \neq 0. \tag{B7}$$

This requires $\Pi P_l|\psi_0\rangle \neq 0$ and $\Pi P_r|\psi_0\rangle \neq 0$. From Eq. (B5), we know this requires $l \geq d$ and $r \geq d$, and it contradicts with $l + r < 2d$. Therefore, we conclude the leading order correction of $\text{Tr}(\Pi\rho_\epsilon^2\Pi O)$ is of order $\mathcal{O}(p^{2d})$.

For higher power of $\Pi\rho_\epsilon^m\Pi$, one may expect the leading order correction will be $\mathcal{O}(p^{md})$. Depending on the particular logical state $|\psi_0\rangle$ and error correction code, the performance may or may not reach $\mathcal{O}(p^{md})$. This is because there can exist shortcuts that make the leading order correction larger than $\mathcal{O}(p^{md})$. In practice, we do witness the performance will be improved with larger m.
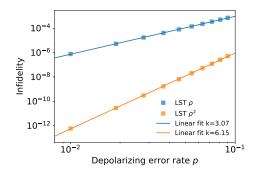
FIG. 5. Infidelity in small error rate region. Theoretically we have shown the leading order correction to infidelity will be $\mathcal{O}(p^{md})$ with $m = 1, 2$. Here, we use $[[5, 1, 3]]$ code with LST as a demonstration. We prepare random logical states and calculate the infidelity. We see the numerical results give linear order correction $\mathcal{O}(p^{3.07})$ and $\mathcal{O}(p^{6.15})$, which is very close to theoretical prediction $\mathcal{O}(p^3)$ and $\mathcal{O}(p^6)$.

Appendix C: 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 closed form expression for $\mathbb{E}[G(P, Q)]$, and $\mathrm{Var}[G(P, Q)]$. Here, we find approximations for the mean and variance using Taylor expansions of $g(P, Q)$.

The approximation for the mean value is

$$
\begin{aligned}
\mathbb{E}[g(P, Q)] &= \mathbb{E}[g(\mu_P, \mu_Q) + g_P'(\mu_P, \mu_Q)(P - \mu_P) + g_Q'(\mu_P, \mu_Q)(Q - \mu_Q) + R] \\
&\approx \mathbb{E}[g(\mu_P, \mu_Q)] + g_P'(\mu_P, \mu_Q)\mathbb{E}[(P - \mu_P)] + g_Q'(\mu_P, \mu_Q)\mathbb{E}[(Q - \mu_Q)] \\
&= g(\mu_P, \mu_Q),
\end{aligned}
\tag{C1}
$$

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

For the variance, we have

$$
\begin{aligned}
\mathrm{Var}[g(P, Q)] &= \mathbb{E}\left\{[g(P, Q) - \mathbb{E}(g(P, Q))]^2\right\} \\
&\approx \mathbb{E}\left\{[g(P, Q) - g(\mu_P, \mu_Q)]^2\right\} \\
&\approx \mathbb{E}\left\{[g_P'(\mu_P, \mu_Q)(P - \mu_P) + g_Q'(\mu_P, \mu_Q)(Q - \mu_Q)]^2\right\} \\
&= g_P'^2(\mu_P, \mu_Q)\,\mathrm{Var}(P) + g_Q'^2(\mu_P, \mu_Q)\,\mathrm{Var}(Q) + 2g_P'(\mu_P, \mu_Q)g_Q'(\mu_P, \mu_Q)\,\mathrm{Cov}(P, Q).
\end{aligned}
\tag{C2}
$$

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

$$
\mathrm{Var}(P/Q) \approx (\frac{\mu_P}{\mu_Q})^2 \left[\frac{\mathrm{Var}(P)}{\mu_P^2} + \frac{\mathrm{Var}(Q)}{\mu_Q^2} - 2\frac{\mathrm{Cov}(P, Q)}{\mu_P \mu_Q} \right]
\tag{C3}
$$

Appendix D: Proof of sample complexities

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

$$
o = \mathrm{Tr}(\rho O).
\tag{D1}
$$

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

$$
\mathbb{E}\hat{o} = \mathrm{Tr}(\rho O).
\tag{D2}
$$

In practice, the expectation $\mathbb{E}\hat{o}_i$ is replaced by a sample mean estimator, $o_{\mathrm{avg}} = \frac{1}{M}\sum_{i=1}^M \hat{o}_i = \frac{1}{M}\sum_{i=1}^M \mathrm{Tr}(O\hat{\rho}_i)$. 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 statistial 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

$$M \geq \text{Var}(\hat{o})/\epsilon\delta^2. \tag{D3}$$

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

$$\begin{aligned}
\text{Var}(\hat{o}) = \mathbb{E}[\hat{o}^2] - \mathbb{E}[\hat{o}]^2 &\leq \mathbb{E}[\hat{o}^2] \\
&= \mathbb{E}_{U\sim\mathcal{U}} \sum_{b\in\{0,1\}^n} \langle b|U\sigma U^\dagger|b\rangle \langle b|U\mathcal{M}^{-1}(O)U^\dagger|b\rangle^2 \\
&\leq ||O||_{\text{shadow}}^2,
\end{aligned} \tag{D4}$$

where the shadow norm of an observable is defined as

$$\begin{aligned}
||O||_{\text{shadow}} &= \max_{\sigma:\text{state}} \left(\mathbb{E}_{U\sim\mathcal{U}} \sum_{b\in\{0,1\}^n} \langle b|U\sigma U^\dagger|b\rangle \langle b|U\mathcal{M}^{-1}(O)U^\dagger|b\rangle^2 \right)^{1/2} \\
&= \max_{\sigma:\text{state}} \left(\mathbb{E}_{U\sim\mathcal{U}} \sum_{b\in\{0,1\}^n} \text{Tr}(\sigma U^\dagger|b\rangle\langle b|U\langle b|U\mathcal{M}^{-1}(O)U^\dagger|b\rangle^2) \right)^{1/2} \\
&= \max_{\sigma:\text{state}} \left(\text{Tr}\, \sigma V_{\mathcal{U}}[O] \right)^{1/2},
\end{aligned} \tag{D5}$$

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

$$\begin{aligned}
V_{\mathcal{U}}[O] &= \mathbb{E}_{U\sim\mathcal{U}} \sum_{b\in\{0,1\}^n} U^\dagger|b\rangle\langle b|U\langle b|U\mathcal{M}^{-1}(O)U^\dagger|b\rangle^2 \\
&= \sum_{b\in\{0,1\}^n} \sum_{\sigma,\tau\in S_3} \text{Wg}[\sigma\tau^{-1}g_0]A[\sigma]B[\tau],
\end{aligned} \tag{D6}$$

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:

$$A[\sigma] = \boxed{\sigma} \quad , \quad B[\tau] = \boxed{\tau} \tag{D7}$$

In the following, we will mainly focus on the analysis of $V_{\mathcal{U}}[O]$ operator and $||O||_{\text{shadow}}^2$. In the main text, we focus on the scheme of encoding each logical qubit with $[n,1]$ stabilizer code, and doing quantum computation with total physical qubits $N = n \times l$, where l is the number of logical qubits. For the classical shadow tomography part, we will use random unitaries sampled from $\mathcal{Cl}(2^n)^{\otimes l}$. One reason of choosing this factorized random unitary group is global clifford group $\mathcal{Cl}(2^{nl})$ is harder to implement in experiments. And the difficulty of implementing this factorized scheme does not depend on number of logical qubits. In practice, it is possible to encode each logical qubit with a small error correction code, such as $[5,1]$ code, and implement random circuits from $\mathcal{Cl}(2^n)$. If the random unitary ensemble is $\mathcal{Cl}(2^n)^{\otimes l}$, then it is easy to show the reconstruction map is

$$\mathcal{M}^{-1}[\sigma] = \otimes_{i=1}^l \left[(2^n + 1)\sigma_i - \text{Tr}(A_i)\mathbb{1} \right], \tag{D8}$$

where σ_i is the reduced classical shadow on part i. The logical Pauli observables will be factorized on each logical sectors, i.e. $O = O_1 \otimes O_2 \otimes \cdots \otimes O_l$. And since the random untaries are sampled from ensemble $\mathcal{Cl}(2^n)^{\otimes l}$, they also have the tensor product structure, i.e. $U = U_1 \otimes U_2 \otimes \cdots \otimes U_l$. By combining those two properties, we can show

$$V_{\mathcal{U}}[O] = \otimes_{i=1}^l V_{\mathcal{U}_i}[O_i]. \tag{D9}$$

Therefore, we only need to focus on the property of $V_{\mathcal{U}_i}[O_i]$ for each logical sector. In the following, we will use $d = 2^n$ for the Hilbert space dimension for one logical sector.

 a. *The calculation for* $V_{\mathcal{U}_i}[P_i\mathbb{1}_i]$: For projection operator P_i, $\mathcal{M}^{-1}(P_i\mathbb{1}_i) = (d+1)P_i - 2\mathbb{1}_i$. And Eq. (D6) can be evaluated

$$V_{\mathcal{Cl}(d)}[P_i\mathbb{1}_i] = \frac{2d-2}{d+2}(P_i + \mathbb{1}_i). \tag{D10}$$

 b. *The calculation for* $V_{\mathcal{U}_i}[P_iO_i]$: For non-trivial Pauli string O_i, $\mathcal{M}^{-1}(P_iO_i) = (d+1)P_iO_i$. And Eq. (D6) can be evaluated

$$V_{\mathcal{Cl}(d)}[P_iO_i] = \frac{2d+2}{d+2}(P_i + \mathbb{1}_i). \tag{D11}$$

As we can see, $V_{\mathcal{Cl}(d)}[P_i\mathbb{1}_i] \lesssim V_{\mathcal{Cl}(d)}[P_iO_i] = \frac{2d+2}{d+2}(P_i+\mathbb{1}_i)$. This result indicates the sample complexity for predicting logical Pauli operators $O = \otimes_{i=1}^{l}O_i$ after projection by $P = \otimes_{i=1}^{l}P_i$ does not depend on the locality of the logical Pauli operators,

$$||PO||_{\text{shadow}}^2 \lessapprox \max_{\sigma:\text{state}} \text{Tr}\left(\sigma \left(\frac{2d+2}{d+2}\right)^l \otimes_{i=1}^{l}(P_i + \mathbb{1}_i)\right) \lesssim 4^l. \tag{D12}$$

This result is different from the sample complexity from local Clifford group or tensored Clifford group $\mathcal{Cl}(d)^{\otimes l}$, where the sample complexity will depends on the locality of Pauli string O. This difference is mainly introduced by the logical subspace projection P. Even the Pauli string O is trivial in some region, the subspace projection P_i will still introduce fluctuation.