I. INTRODUCTION

Learning the properties of an unknown but physically accessible quantum system is a fundamental task in quantum information processing [1]. A standard tool for this task is quantum tomography, a process by which one recovers a classical description of a quantum system through performing measurements on it. Unfortunately, finding the full description of a quantum system by quantum tomography is computationally-intensive; it requires an exponential number of copies of the system [2–4].

Recently, Huang, Kueng, and Preskill introduced a novel method—the classical shadow paradigm [1]—to circumvent the above limitation. A key insight behind classical shadows rests on the fact that in many cases one does not need to learn a complete description of a quantum system; one can learn its most useful properties from a minimal sketch of the quantum system, the classical shadow. The performance of the classical shadow protocol depends on several factors. In particular, it depends on an ensemble of unitary operators from which an operation is chosen randomly to be applied to the unknown quantum state. A user must choose this ensemble in advance, according to some desiderata, such as the need for the shadow channel to be invertible and for one to have an efficient algorithm for sampling a unitary from the ensemble. Several unitary ensembles have been considered by previous authors, including the local and global Clifford ensembles [1], fermionic Gaussian unitaries [5], chaotic Hamiltonian evolutions [6], and locally scrambled unitary ensembles [7].

One can ask: what is the weakest assumption on the unitary ensemble that would still yield meaningful and interesting results? Our candidate solution to this question is the assumption that the unitary ensemble is invariant under multiplication by a Pauli operator. Ensembles satisfying this assumption—namely the Pauli-invariant unitary ensembles—include a wide range of ensembles, including the Pauli group and the aforementioned local and global Clifford group and locally scrambled unitary ensembles.

Let us summarize our main contributions in this work: For Pauli-invariant unitary ensembles, we find an explicit formula for the reconstruction map corresponding to the shadow channel. We establish a connection between the coefficients of the reconstruction map and the entanglement features of the dynamics. We give upper bounds on the sample complexity in terms of the shadow norm for the task of expectation estimation using the classical shadow protocol.

We also give two applications of our methods. First, we apply our results to locally scrambled unitary ensembles, giving the reconstruction map explicitly, along with sample complexity bounds. Unlike the approach taken in [7], our study circumvents the need to solve an exponential-sized system of linear equations. Second, we apply our results to the shadow process tomography of quantum channels, where we generalize the results of [8, 9] to allow for any Pauli-invariant unitary ensemble.

Related work.—The classical shadow paradigm, with the sample efficiency it touts, has attracted considerable attention over the last couple of years [10]. Several applications have been proposed, including applying the classical shadow framework to estimate expectation values of molecular Hamiltonians [11, 12], detecting or estimating the degree of entanglement in quantum systems [1, 13–15], classifying quantum data [16], measuring out-of-time-ordered correlators [17], approximating wave function overlaps [18], solving quantum many-body problems [1, 19, 20], estimating gate-set properties [21], and avoiding barren plateaus [22]. Noise analyses
of the classical shadow protocol have been performed, with noise-resilient versions developed [23–25].

Authors have proposed variants and generalizations of the protocol, including locally-biased classical shadows [11], derandomization techniques [26], decision diagram techniques [27], grouping strategies [28], PovM-based classical shadows [29], classical shadows with locally-scrambled unitary ensembles [7], and classical shadows with unitary ensembles generated by a Hamiltonian [6]. Extensions have been found to fermions [5] and to quantum channels [8, 9]. With respect to certain figures-of-merit, the protocol has been shown to give an advantage over alternative approaches [11, 30] and has been analyzed theoretically from a Bayesian point of view [31]. Lower bounds on the performance of classical shadows have also been proved [1, 32].

II. BACKGROUND: CLASSICAL SHADOWS

Let us begin by describing the classical shadow protocol (see also Figure 1, where we present a circuit diagram for a single round of the protocol) that was introduced in [1]. The input of the protocol is an unknown but physically accessible quantum state $\rho$ and the output of the protocol is a set of classical shadows $\{\hat{\rho}_i\}$ that are unbiased estimators of $\rho$, i.e. the expectation value of each $\hat{\rho}_i$ is precisely $\rho$. In an estimation task, these classical shadows may subsequently be used to estimate expectation values of given quantum observables in the state $\rho$.

Before the protocol is run, a user will have to pre-decide on a unitary ensemble $\mathcal{E} = \{ (U, P(U)) \}_{U \in \mathcal{U}}$, where $\mathcal{U}$ is a set of unitary operators and $P$ is a probability distribution on $\mathcal{U}$. The first step of the protocol involves choosing a random unitary $U$ from the unitary ensemble $\mathcal{E}$ according to the specified distribution $P(U)$. The unitary $U$ is then applied to $\rho$ and a computational basis measurement is performed on the resultant state to obtain an $n$-bit string $\hat{b} = b_1 b_2 \ldots b_n \in \{0, 1\}^n$. The random unitary $U$ and random string $\hat{b}$ obtained are then combined to give the state

$$\hat{\sigma}_{U,\hat{b}} = U^\dagger |\hat{b}\rangle \langle \hat{b}| U,$$

which is stored in classical memory. Note that this produces an ensemble of states

$$\mathcal{E}_\rho = \left\{ \left( \hat{\sigma}_{U,\hat{b}}, P(U, \hat{b}) \right) \right\}_{U,\hat{b}}$$

where $P(U, \hat{b}) = P(U) P(\hat{b}|U)$ and $P(\hat{b}|U) = \text{Tr} \left[ \hat{\sigma}_{U,\hat{b}} \rho \right]$, as given by the Born rule. The expected value of this ensemble is given by

$$\sigma = \mathbb{E}_{\hat{b} \in \mathcal{E}_\rho} \hat{\sigma} = \mathbb{E}_{\hat{b}} \sum_{U} \hat{\sigma}_{U,\hat{b}} \text{Tr} \left[ \hat{\sigma}_{U,\hat{b}} \rho \right] := \mathcal{M}[\rho],$$

where $\mathcal{M}$—called the shadow channel—is a completely positive and trace-preserving map. To construct the classical shadow, the shadow channel needs to be invertible. If the conditions for invertibility are met, the inverse of the shadow channel $\mathcal{M}^{-1}$, called the reconstruction map, is applied to the classically stored $\hat{\sigma}_{U,\hat{b}}$ to obtain the classical snapshot

$$\hat{\rho} = \mathcal{M}^{-1}(\hat{\sigma}_{U,\hat{b}}),$$

which is called the classical shadow. As required, $\hat{\rho}$ is an unbiased estimator of $\rho$:

$$\rho = \mathcal{M}^{-1}[\sigma] = \mathbb{E}_{\hat{\sigma} \in \mathcal{E}_\rho} \mathcal{M}^{-1}[\hat{\sigma}] = \mathbb{E}_{\hat{b}, U} \mathcal{M}^{-1}(\hat{\sigma}_{U,\hat{b}}) = \mathbb{E}_{\hat{b}, U} [\hat{\rho}].$$

Note that while $\mathcal{M}^{-1}$ is trace-preserving, it is not necessarily completely positive (this is fine since $\hat{\rho}$ does not need to be a quantum state). One repeats this process multiple times to construct an ensemble $\{\hat{\rho}_i\}$ of classical shadows. The number of times this process needs to be repeated—which translates to the number of samples of $\hat{\rho}$ needed—depends on the number $N$ of different properties of $\rho$ that one wants to estimate. Since quantum measurements are destructive, one might think naively that one would need at least $N$ samples, where at least one sample is used for each property. Surprisingly, with classical shadows, an order of $\log N$ samples is sufficient [1].

As illustrated above, the reconstruction map $\mathcal{M}^{-1}$ plays a pivotal role in the classical shadow protocol. However, for arbitrary unitary ensembles, no general closed-form analytic formula is known for $\mathcal{M}^{-1}$. Instead, prior to this study, it was only for a handful of unitary ensembles (e.g. the local and global Clifford ensembles) that analytic expressions have been derived. In this study, We address this gap by presenting an explicit formula for the reconstruction map for the wide-class of Pauli-invariant unitary ensembles, thereby enlarging the class of ensembles for which such expressions are known.

III. MAIN RESULTS

A. Noiseless classical shadows

Let us now introduce the Pauli-invariant unitary ensembles, and derive results for the performance of classical shadow protocols that utilize such ensembles. For simplicity, we shall start by assuming that the classical shadow protocol is implemented perfectly without any noise. This assumption will subsequently be relaxed in Section III B.
Pauli-invariant unitary ensemble—We denote the set of Pauli operators on $n$ qubits by $\mathcal{P}_n = \{ P_i = \otimes_j P_{a_j} : a \in V^n \}$, where $V := Z_2 \times Z_2 = \{ (0,0), (0,1), (1,0), (1,1) \}$, and $P_{(0,0)} = I$, $P_{(0,1)} = Z$, $P_{(1,0)} = X$, $P_{(1,1)} = Y$. Given a unitary ensemble $\mathcal{E} = \{ (U, P(U)) \}_{\mathcal{U}}$, $\mathcal{E}$ is called Pauli-invariant if the probability distribution $P(U)$ satisfies the following (right) Pauli-invariant condition, that is

$$P(U) = P(U P_{a}), \quad \forall P_{a} \in \mathcal{P}_n. \quad (5)$$

This is a weak assumption, and many unitary ensembles satisfy this condition, for example, (1) the Pauli group $\{ \pm 1, \pm i \} \times \mathcal{P}_n$; (2) $D$-dimensional local, random quantum circuits; (3) local Clifford unitaries; and (4) global Clifford unitaries.

For Pauli-invariant unitary ensembles, we will now provide an explicit formula for the shadow channel $\mathcal{M}$ and the corresponding reconstruction map $\mathcal{M}^{-1}$, if it exists. For clarity of presentation, the detailed proofs of all the results in this subsection are provided in Appendix B.

**Theorem 1.** For a Pauli-invariant unitary ensemble $\mathcal{E}$, the shadow channel can be written as

$$\mathcal{M}[\rho] = \frac{1}{2^n} \sum_{a \in V^n} W_{\mathcal{E}}[\vec{a}] \Tr[\rho P_{a}] P_{a}, \quad (6)$$

where $W_{\mathcal{E}}[\vec{a}]$ is the average squared Pauli coefficient of the classical shadow, defined as

$$W_{\mathcal{E}}[\vec{a}] = E_{h} E_{l} W_{\mathcal{E}_{h,l}}[\vec{a}], \quad (7)$$

where $E_{h} := \frac{1}{2^n} \sum_{h}$ denotes the average with respect to the uniform distribution over $n$-bit strings. For a state $\sigma$, we have used the notation $W_{\mathcal{E}}[\vec{a}] := |\Tr[\sigma P_{a}]|^2$ to denote the square of the $\vec{a}$-th Pauli coefficient of $\sigma$.

For Pauli-invariant unitary ensembles, the reconstruction map $\mathcal{M}^{-1}$ exists if and only if $W_{\mathcal{E}}[\vec{a}] > 0$ for all $\vec{a}$. If it exists, the reconstruction map is completely specified by its action on the Pauli basis elements as follows:

$$\mathcal{M}^{-1}[P_{a}] = W_{\mathcal{E}}[\vec{a}]^{-1} P_{a}. \quad (8)$$

Now, let us discuss the connection between the coefficients of the reconstruction map $W_{\mathcal{E}}$ and the 2nd entanglement feature [34, 35], which has been used to describe the entangling power of unitary ensembles and is defined as follows:

$$E_{\mathcal{E}}^{(2)}[A] = E_{l} E_{k} e^{-\sigma_{A}(\delta_{v,l})}, \quad (9)$$

where $A$ is a subset of $[n]$, $S_{\mathcal{E}}^{(2)}(\sigma) = -\log \Tr[\sigma_{A}]$, where $\sigma_{A} = \Tr_{[n]}[\sigma]$ denotes the 2nd Rényi entanglement entropy of the state $\sigma$ on the subset $A$. For any subset $S \subset [n]$, let us define $W_{\mathcal{E}}[S]$ as the sum of the coefficients whose support is $S$, i.e.,

$$W_{\mathcal{E}}[S] = \sum_{\vec{a} : \text{supp}(\vec{a}) = S} W_{\mathcal{E}}[\vec{a}], \quad (10)$$

where $\text{supp}(\vec{a})$ denotes the support of the vector $\vec{a}$. The following proposition expresses $W_{\mathcal{E}}[S]$ in terms of the 2nd entanglement feature.

**Proposition 2.** The coefficients of the reconstruction map can be expressed as follows:

$$W_{\mathcal{E}}[S] = (-1)^{|S|} \sum_{A \subset [S]} (-2)^{|A|} E_{\mathcal{E}}^{(2)}[A]. \quad (11)$$

Now, let us consider the sample complexity for the task of expectation estimation using classical shadows with Pauli-invariant unitary ensembles. We shall make use of the result from [1] that the sample complexity of classical shadows with the input state $\rho$ and an observable $O$ is upper-bounded by the following (squared) shadow norm:

$$\|O\|_{\mathcal{E}}^2 = E_{\mathcal{E}} \delta[\delta]^2, \quad (12)$$

where the estimator $\hat{\delta}$ of the observable $O$ is given by $\hat{\delta} = \Tr[O \mathcal{M}^{-1}[\delta]]$. More precisely, the sample complexity $\mathcal{S}$ needed to accurately predict a collection of $N$ linear target functions $\{ \Tr[O\rho] \}_{i=1}^N$ with error $\epsilon$ and failure probability $\delta$ is [1]

$$\mathcal{S} = O \left( \frac{\log(N/\delta)}{\epsilon^2} \max_{1 \leq i \leq N} \left\| O_i - \frac{1}{2^n} \Tr[O_i] \right\|_{\mathcal{E}}^2 \right).$$

In general, the shadow norm is hard to compute, though for special cases, closed-form expressions for the shadow norm can be derived. To this end, let us first consider the special but widely-used example wherein the observable $O$ is taken to be a Pauli operator $P_{a}$.

**Proposition 3.** If the observable $O$ is the Pauli operator $P_{a}$, then the shadow norm is equal to

$$\|P_{a}\|_{\mathcal{E}}^2 = W_{\mathcal{E}}[\vec{a}]^{-1}. \quad (13)$$

Note that the shadow norm $\|O\|_{\mathcal{E}}$ usually depends on the input state $\rho$. If we interested in the expectation of the shadow norm over some unitary ensemble instead of some specific state, we will need a notion of an average shadow norm. To this end, let us consider the (squared) average shadow norm over the Pauli group, defined as $\|O\|_{\mathcal{E}}^2 = E_{\mathcal{V} \in \mathcal{P}_n} \|O\|_{\mathcal{E}_{\mathcal{V}P_{a}}}^2$. The following proposition gives an expression for the average shadow norm.

**Proposition 4.** The (squared) average shadow norm over the Pauli group $\|O\|_{\mathcal{E}}^2$ can be expressed as follows:

$$\|O\|_{\mathcal{E}}^2 = \frac{1}{4^n} \sum_{a \in V^n} W_{\mathcal{E}}[\vec{a}]^{-1} W_{O}[\vec{a}], \quad (14)$$

where $W_{O}[\vec{a}] = |\Tr[O P_{a}]|^2$.

Therefore, given a set of $N$ traceless observables $\{ O_i \}_{i=1}^N$, the average shadow norm provides the following lower bound for the sample complexity:

$$\mathcal{S} \geq \frac{\log(N/\delta)}{\epsilon^2} \max_{1 \leq i \leq N} \langle \hat{W}_{\mathcal{E}}^{-1}, \hat{W}_{O_i} \rangle, \quad (15)$$

where the (normalized) inner product is defined as $\langle \hat{W}_{\mathcal{E}}^{-1}, \hat{W}_{O_i} \rangle := \frac{1}{4^n} \sum_{a \in V^n} W_{\mathcal{E}}[\vec{a}]^{-1} W_{O_i}[\vec{a}]$. 
B. Noisy classical shadows

Our assumption thus far has been that the classical shadow protocol is implemented perfectly without being affected by noise. However, this is an unrealistic scenario, as noise is unavoidable in real-world experiments. For classical shadows to be useful even in the presence of noise, it is necessary to employ error mitigation techniques. This is especially important in the present noisy intermediate-scale quantum era [36, 37], where quantum devices are noisy and quantum algorithms are performed without quantum error-correction [38]. Fortunately, as previous works have shown [23, 24], the classical shadow protocol can be made noise-resilient by making modifications to the shadow channel.

In this subsection, let us consider noisy classical shadows with a Pauli-invariant unitary ensemble. Similar to the setting in [23, 24], we shall assume that a noise channel $\Lambda$ acts on the pre-measurement state $U\rho U^\dagger$ just before the measurement is performed. Such an assumption is obeyed by gate-independent, time-stationary, and Markovian noise [23, 39]. In this case, the classical shadow is still defined as $\mathcal{M}^{-1}([\hat{\sigma}_{U,b}])$, although the probability distribution is now changed to $P_\Lambda(U, b) = P(U)P_{a}(b|U)$ where $P_{a}(b|U) = \text{Tr}[b|b\Lambda[U\rho U^\dagger]]$. Similar to the noiseless case (2), the ensemble of states is given by

$$E_{\Lambda,\rho} = \left\{ \left( \hat{\sigma}_{U,b}, P(U, b) \right) \right\}_{U,b}.$$ 

Hence, by taking the average of the post-measurement states, we have $\sigma = \mathbb{E}_{U} \sum_{b} \hat{\sigma}_{U,b} \text{Tr} \left[ b|b\Lambda[U\rho U^\dagger] \right] = \mathcal{M}_{\Lambda}[^1\sigma]$, where $\mathcal{M}_{\Lambda}$ is the noisy channel, and $\rho = \mathcal{M}^{-1}[\sigma] = \mathbb{E}_{\hat{\sigma} \in E_{\Lambda}} \mathcal{M}^{-1}_{\Lambda}[\hat{\sigma}]$, where $\mathcal{M}^{-1}_{\Lambda}$ is the noisy reconstruction map. We will now provide an explicit form for the noisy shadow channel $\mathcal{M}_{\Lambda}$ and the noisy reconstruction map $\mathcal{M}^{-1}_{\Lambda}$. For clarity of presentation, the detailed proof of all the results in this subsection are provided in Appendix C.

**Theorem 5.** Given a Pauli-invariant unitary ensemble $E$ and a noise channel $\Lambda$, the noisy shadow channel is given by

$$\mathcal{M}_{\Lambda}[\rho] = \frac{1}{2^n} \sum_{d} W_{E_{\Lambda}}[d] \text{Tr}[\rho P_{d}^\dagger] P_{d},$$

where $W_{E_{\Lambda}}[\bar{d}]$ is the average Pauli coefficient of the noisy classical shadow, and is defined as

$$W_{E_{\Lambda}}[\bar{d}] = \mathbb{E}_{b} \mathbb{E}_{U} \text{Tr} \left[ \hat{\sigma}_{U,b} P_{\bar{d}} \right] \text{Tr} \left[ U^\dagger \Lambda^d [b|b] U P_{\bar{d}} \right].$$

For the Pauli-invariant unitary ensemble, $\mathcal{M}^{-1}_{\Lambda}$ exists if and only if $W_{E_{\Lambda}}[\bar{d}] > 0$ for all $\bar{d}$. If it exists, the reconstruction map is defined by

$$\mathcal{M}^{-1}_{\Lambda}[P_{\bar{d}}] = W_{E_{\Lambda}}[\bar{d}]^{-1} P_{\bar{d}}.$$  

Now, let us consider the sample complexity of the noisy classical shadow protocol. Similar to the noiseless case, the sample complexity of the noisy classical shadow with the input state $\rho$ and observable $O$ is upper bounded by the (squared) noisy shadow norm $\|O\|_{E_{\Lambda},\rho}^2 = \mathbb{E}_{\hat{\sigma} \in E_{\Lambda}} \hat{\sigma}^2$, where the estimator of the observable is taken to be $\hat{\sigma} = \text{Tr}[O\mathcal{M}_{\Lambda}^{-1}[\hat{\sigma}]]$.

**Proposition 6.** If the observable $O$ is taken to be the Pauli operator $P_{\bar{d}}$, the shadow norm is equal to

$$\|P_{\bar{d}}\|_{E_{\Lambda},\rho}^2 = W_{E_{\Lambda}}[\bar{d}]^{-2} W_{E_{\Lambda}}[\bar{d}],$$

where $W_{E_{\Lambda}}[\bar{d}]$ is defined as

$$W_{E_{\Lambda}}[\bar{d}] = \mathbb{E}_{U} \mathbb{E}_{b} \text{Tr} \left[ P_{\bar{d}} \hat{\sigma}_{U,b} \right] \text{Tr} \left[ [b|b][b|b] \right].$$

Hence, if $\Lambda$ is unital, then $W_{E_{\Lambda}}[\bar{d}] = W_{E}[\bar{d}]$.

From the above proposition, we note that $W_{E_{\Lambda}}[\bar{d}]$ depends on only how the noise channel $\Lambda$ acts on the identity operator. In the case where $\Lambda$ is unital, $W_{E_{\Lambda}}[\bar{d}]$ reduces to $W_{E}[\bar{d}]$, the analogous quantity in the noiseless case, which suggests a particular role of noise unitality in the analysis of classical shadows.

Similar to the noiseless case, let us define the (squared) average shadow norm over the Pauli group for the noisy classical shadow to be $\|O\|_{E_{\Lambda}}^2 = \mathbb{E}_{\bar{d} \in P_{n}} \|O\|_{E_{\Lambda},\rho}^2$. The following proposition gives an explicit expression for the average shadow norm.

**Proposition 7.** The average shadow norm in the noisy classical shadow protocol with the noise channel $\Lambda$ can be expressed as follows

$$\|O\|_{E_{\Lambda}}^2 = \frac{1}{4^n} \sum_{\bar{d}} W_{E_{\Lambda}}[\bar{d}]^{-2} W_{E}[\bar{d}] W_{O}[\bar{d}].$$

Let us now provide an example of noisy classical shadows and demonstrate how the performance of the protocol depends on the noise rate.

**Example 1** (Noisy classical shadows with global depolarizing noise). Let us consider the case where the noise channel is the global depolarizing channel, defined as $D_{p}[\cdot] = (1-p)[\cdot] + p\text{Tr}[\cdot]I$. Since $D_{p}$ is unital, we have $W_{E_{D_{p}}}[\bar{d}] = W_{E}[\bar{d}]$. Hence, the coefficients of the shadow channel may be expressed as

$$W_{E_{D_{p}}}[\bar{d}] = (1-p)^{-\delta_{\bar{d},\bar{0}}} W_{E}[\bar{d}].$$

Thus, the shadow norm for a Pauli operator $P_{\bar{d}}$ obeys the following identity

$$\|P_{\bar{d}}\|_{E_{D_{p}},\rho}^2 = (1-p)^{2\delta_{\bar{d},\bar{0}}-2} \|P_{\bar{d}}\|_{E,\rho}^2,$$

where $\delta_{\bar{d},\bar{0}}$ denotes the Kronecker delta function. From the above equation, we see that depolarizing noise increases the number of samples needed for expectation estimation, with an increase that is proportional to the noise rate.
IV. APPLICATION OF OUR RESULTS

A. Classical shadow with locally scrambled unitary ensembles

First, let us apply our results to locally scrambled unitary ensembles. A unitary ensemble is said to be locally scrambled [7] if the probability distribution $P(U)$ satisfies local basis invariance, that is,

$$P(U) = P(UV), \quad \forall \text{ unitaries } V = V_1 \otimes \cdots \otimes V_n.$$  \hspace{1cm} (24)

It is easy to see that the Pauli-invariance assumption is weaker than the locally scrambled assumption, as locally scrambled unitary ensembles are Pauli-invariant. Not all Pauli-invariant unitary ensembles are locally scrambled though—a counterexample is the Pauli group.

Classical shadows with locally scrambled unitary ensembles were previously considered in [7], which expressed the reconstruction map as the solution of a linear system of size $O(2^n)$; an explicit formula for the reconstruction map was not provided. In our study, however, since locally scrambled unitary ensembles are a special case of the Pauli-invariant unitary ensembles, a consequence of Theorem 1 is an explicit formula for the reconstruction map that circumvents the need to solve an exponential-sized linear system. Our next proposition, whose proof we provide in Appendix D, makes this explicit.

**Proposition 8.** Given a locally scrambled unitary ensemble $\mathcal{E}$, the shadow channel is

$$\mathcal{M}[\rho] = \frac{1}{2^n} \sum_{S \subseteq [n]} \bar{W}_E[S] \sum_{d : \text{supp}(d) = S} \text{Tr}[\rho P_d] P_d,$$

where $\bar{W}_E[S]$ is defined as $\bar{W}_E[S] = \mathbb{E}_{\text{d : supp}(d) = S} W_E[d]$, and the reconstruction map is given by

$$\mathcal{M}^{-1}[P_d] = \bar{W}_E[\text{supp}(d)]^{-1} P_d.$$

Our expression for the reconstruction map above may be compared with that given in [7], which expressed the reconstruction map as $\mathcal{M}^{-1}[\sigma] = \sum_{S \subseteq [n]} r_S D^S[\sigma]$, where $D^S$ denotes the $|S|$-fold Kronecker product of the single-qubit erasure channel $D[\cdot] = \text{Tr} [ \cdot I/2 ]$ with itself acting on all the qubits indexed by $S$, and the coefficients $r_S$ are given as the solution of a linear system whose coefficients are the entanglement features. In the next proposition, we find an explicit formula for $r_S$ in terms of the coefficients $\bar{W}_E[S]$.

**Proposition 9.** Given a reconstruction map written as $\mathcal{M}^{-1}[\sigma] = \sum_S r_S D^S[\sigma]$, the coefficients $r_S$ can be expressed in terms of $\bar{W}_E[A]$ as follows

$$r_S = \sum_{A \subseteq S} (-1)^{|S| - |A|} \bar{W}_E[A]^{-1}, \hspace{1cm} (25)$$

where $A^c$ denotes the complement of $A$ in $[n]$.

The proof of Proposition 9 is presented in Appendix D. Using Proposition 2 and the fact that $W_E[S] = \sum S \bar{W}_E[S]$, our next corollary shows how the coefficients $r_S$ can be expressed in terms of the entanglement feature.

**Corollary 10.** Given a reconstruction map written as $\mathcal{M}^{-1}[\sigma] = \sum_S r_S D^S[\sigma]$, the coefficients $r_S$ can be expressed as follows

$$r_S = (-1)^{|S|} \sum_{A \subseteq S} \left( \sum_{B \subseteq A^c} (-2)^{|B|} E_E[B] \right)^{-1} \sum_{A \subseteq S} (-2)^{|A|} E_E[A].$$  \hspace{1cm} (26)

For the locally scrambled unitary ensembles, the average shadow norm is defined in [7] as $\|O\|_E^2 = \mathbb{E}_{\rho \in \mathcal{U}(2^n)} \|O\|_E^2$, which provides a typical lower bound for the sample complexity of expectation estimation using classical shadows with locally scrambled unitary ensembles.

Just as we have provided an explicit formula for the reconstruction map in terms of the entanglement feature, so too can we provide an explicit formula for the shadow norm using the entanglement feature.

**Proposition 11.** Given a locally scrambled unitary ensemble $\mathcal{E}$, the average shadow norm is

$$\|O\|_E^2 = \frac{1}{4^n} \sum_{S \subseteq [n]} \bar{W}_E[S]^{-1} W_O[S],$$  \hspace{1cm} (27)

where $W_O[S] = \sum_{d : \text{supp}(d) = S} W_O[d].$

The proof of Proposition 11 is presented in Appendix D. Based on Propositions 2 and 11, we can express the average shadow norm in terms of the entanglement feature as follows:

$$\|O\|_E^2 = \frac{1}{4^n} \sum_{S \subseteq [n]} (-3)^{|S|} W_O[S] \left( \sum_{A \subseteq S} (-2)^{|A|} E_E[A] \right)^{-1}.$$  \hspace{1cm} (28)

In summary, this section saw an application of our results to classical shadows with locally scrambled unitary ensembles, where we obtained an explicit formula for the reconstruction map and the average shadow norm in terms of the entanglement features, thus circumventing the need to solve the exponential-sized system of linear equations in [7]. These explicit formulae may be helpful for the further analysis of the role of entanglement in the classical shadow protocol, which we shall leave for future work.

B. Classical shadow tomography for quantum channels with Pauli-invariant unitary ensembles

The classical shadow paradigm was recently extended to the shadow tomography of quantum channels [8, 9]. In these works, the unitary ensembles considered were the local and global Clifford ensembles. In this study, we extend their results to the case of Pauli-invariant unitary ensembles.

Next, we shall outline the procedure proposed in [8, 9] for constructing the classical shadows for quantum channels. The unitary ensembles used will be assumed to be Pauli-invariant.
(1) Prepare $\tilde{b}_i$ with $\tilde{b}_i \in \{0,1\}^n$ chosen uniformly randomly.
(2) Apply a unitary $U_i$ chosen from the locally Pauli-invariant ensemble $\mathcal{E}_i$.
(3) Apply the quantum channel $\mathcal{T}$.
(4) Apply a unitary $U$ chosen from the locally Pauli-invariant ensemble $\mathcal{E}_o$, where the unitary ensemble $\mathcal{E}_o$ may be different from $\mathcal{E}_i$.
(5) Measure in the Pauli $Z$ basis to get the output $\tilde{b}_o \in \{0,1\}^n$.

The post-measurement state is given by
$$\hat{\sigma}_{\bar{i},o} = U_T^\dagger \hat{b}_i[b_i]U_i^* \otimes U_T^\dagger \hat{b}_o[b_o]U_o = \hat{\sigma}_i \otimes \hat{\sigma}_o \quad \text{where} \quad \hat{\sigma}_i = U_T^\dagger \hat{b}_i[b_i]U_i^* \quad \text{and} \quad \hat{\sigma}_o = U_T^\dagger \hat{b}_o[b_o]U_o.$$ Hence, given $\tilde{b}_i,U_i,U_o$, the probability of getting the outcome $\tilde{b}_o$ is given by $P(\tilde{b}_o|\tilde{b}_i,U_i,U_o) = 2^n \text{Tr}[\hat{\sigma}_{\bar{i},o} \otimes \mathcal{J}(T)]$, where $\mathcal{J}(T) = \mathbb{I} \otimes \lambda(\langle \Psi'|\Psi \rangle)$ is the Choi-Jamiołkowski state of $\mathcal{T}$, where $\langle \Psi'|\Psi \rangle = \frac{1}{2^n} \sum_{|i\rangle} |i\rangle \langle i|$ is the Bell state. It is easy to verify that $\sum_{b_o} P(\tilde{b}_o|\tilde{b}_i,U_i,U_o) = 2^n \text{Tr}[\hat{\sigma}_i \otimes \mathcal{J}(T)] = 1$.

Also, the probability of obtaining $\tilde{b}_i,U_i,U_o$ is equal to $P(\tilde{b}_i,U_i,U_o) = P(\tilde{b}_i)P(U_i)P(U_o)$. Hence, the ensemble of states is described by
$$\mathcal{E}_{i,o} = \{ (\hat{\sigma}_{\bar{i},o}, P(\tilde{b}_i,U_i,U_o, \tilde{b}_o) ) \}, \quad (29)$$
where the probability distribution $P(\tilde{b}_i,U_i,U_o, \tilde{b}_o) = P(\tilde{b}_i|U_i,U_o)P(\tilde{b}_o|U_i,U_o)$. Taking the average of the classical shadow, we obtain $\sigma = \mathbb{E}_{\mathcal{E}_o \in \mathcal{E}_o} \hat{\sigma} = \mathcal{M}_{\mathcal{T}} \mathcal{J}(T)$.

To implement classical shadow tomography, the inverse of the shadow channel $\mathcal{M}_{\mathcal{T}}^{-1}$ would need to be implemented. If the inverse exists, then $\mathcal{J}(T) = \mathcal{M}_{\mathcal{T}}^{-1}[\sigma] = \mathcal{E}_{\bar{i},o} \mathcal{M}_{\mathcal{T}}^{-1} [\hat{\sigma}]$. Given a quantum state $\rho$ and an observable $O$, the estimator $\hat{\sigma}$ is defined as
$$\hat{\sigma} = \text{Tr}[\mathcal{J}^T \rho \otimes O].$$ Therefore,
$$\mathcal{E}_{\bar{i},o} = \{ (\mathcal{M}_{\mathcal{T}}^{-1}[\hat{\sigma}], \sigma_o) \rho \otimes O \}. \quad (30)$$

**Proposition 12.** Given two Pauli-invariant unitary ensembles $\mathcal{E}_i$ and $\mathcal{E}_o$, the shadow channel $\mathcal{M}_{\mathcal{T}}$ is
$$\mathcal{M}_{\mathcal{T}}[P_{\tilde{b}_i} \otimes P_{\tilde{b}_o}] = W_{\mathcal{E}_i}[P_{\tilde{b}_i}] W_{\mathcal{E}_o}[P_{\tilde{b}_o}] P_{\tilde{b}_i} \otimes P_{\tilde{b}_o}. \quad (31)$$

The proof of Proposition 12 is presented in Appendix E.

Now, for the classical shadow tomography for a quantum channel, the sample complexity is upper bounded by the following shadow norm:
$$\| \rho^T \otimes O \|_{\mathcal{E}_{\bar{T}}}^2 = \mathbb{E}_{\mathcal{E}_{\bar{T}}} \hat{\sigma}[\hat{\sigma}]^2.$$ Next, let us consider the shadow norm in the case where the observable is some Pauli operator and apply this result to the problem of estimating Pauli channels.

**Proposition 13.** If the observable is taken to be a Pauli operator $P_{\tilde{b}}$, then the shadow norm is equal to
$$\| \rho^T \otimes P_{\tilde{b}} \|_{\mathcal{E}_{\bar{T}}}^2 = W_{\mathcal{E}_o}[\tilde{b}]^{-1} \sum_{\tilde{a}} W_{\mathcal{E}_i}[\tilde{a}]^{-1} W_{\rho}[\tilde{a}]^{-1}. \quad (32)$$

The proof of Proposition 13 is presented in Appendix E.

Example 2 (Estimation of Pauli channels). A quantum channel $\mathcal{T}$ is called a Pauli channel if it can be written as $\mathcal{T}[\cdot] = \sum_{\tilde{d} \in \mathbb{V}_n} p_{\tilde{d}} P_{\tilde{d}}$ with $\sum_{\tilde{d}} p_{\tilde{d}} = 1$. Equivalently, Pauli channels $\mathcal{T}$ are precisely those that can be written as $\mathcal{T}[\cdot] = \sum_{\tilde{d}} \lambda_{\tilde{d}} \text{Tr}[\cdot P_{\tilde{d}}] P_{\tilde{d}}$. It is easy to see that the probability distribution $\{ p_{\tilde{d}} \}$ and the coefficients $\{ \lambda_{\tilde{d}} \}$ are related as follows $\lambda_{\tilde{d}} = \frac{1}{2^n} \sum_{\tilde{d}} p_{\tilde{d}} (1 - |\tilde{d}|, b_{\tilde{d}})_{s} \{ \langle d|b_{\tilde{d}} \rangle \}_{s}$, where $\langle d|b_{\tilde{d}} \rangle_{s} = \langle \tilde{x}_1, \tilde{x}_2, \tilde{y}_1, \tilde{y}_2 | \tilde{y}_1, \tilde{y}_2 \rangle$ for any $\tilde{x}_1 = (x_1, x_2)$, $\tilde{y}_2 = (y_1, y_2) \in \mathbb{V}$. The task of estimating the coefficients $\{ \lambda_{\tilde{d}} \}_b$ of the Pauli channel has been investigated in [39–47]. Here, we consider the classical shadow protocol for estimating coefficients $\lambda_{\tilde{d}}$ with Pauli-invariant unitary ensembles $\mathcal{E}_i$ and $\mathcal{E}_o$. The classical shadow is taken to be $\{ \mathcal{M}_{\mathcal{T}}^{-1}[\hat{\sigma}_{\bar{i},o}] \}$, and the reconstruction map is given in Proposition 12. To estimate the coefficients $\hat{\lambda}_{\tilde{d}}$, let us consider the observable $\mathcal{P}_{\tilde{d}} \otimes \mathcal{P}_{\tilde{d}}$, and take the estimator of the observable to be $\hat{\lambda}_{\tilde{d}} = \text{Tr} [\mathcal{M}_{\mathcal{T}}^{-1}[\hat{\sigma}_{\bar{i},o}] \mathcal{P}_{\tilde{d}} \otimes \mathcal{P}_{\tilde{d}}]$, where it is easy to verify that $W_{\mathcal{E}_{\bar{i},o}}[\mathcal{P}_{\tilde{d}}] = \text{Tr} [\mathcal{J}(\mathcal{T}) \mathcal{P}_{\tilde{d}} \otimes \mathcal{P}_{\tilde{d}}] = \lambda_{\tilde{d}}$. Then the sample complexity $\mathcal{S}$ needed to accurately predict a collection of $4^n$ linear target functions $\{ \mathcal{J}(\mathcal{T}) \mathcal{P}_{\tilde{d}} \otimes \mathcal{P}_{\tilde{d}} \}_{\tilde{d} \in \mathbb{V}_n}$ with error $\varepsilon$ and failure probability $\delta$ is
$$\mathcal{S} = O \left( \frac{n \log(1/\delta)}{\varepsilon^2} \max_{\tilde{d}} W_{\mathcal{E}_o}[\tilde{d}]^{-1} W_{\mathcal{E}_i}[\tilde{d}]^{-1} \right).$$

**V. CONCLUSION**

In this study, we investigated the classical shadow protocol with Pauli-invariant unitary ensembles. First, we provided an explicit formula for the reconstruction map corresponding to the shadow channel and established a connection between the coefficients of the reconstruction map and the entanglement features of the dynamics. Using the shadow norm, we gave explicit sample complexity upper bounds for the estimation task that the classical shadow protocol solves. Finally, we presented two applications of our results. First, we applied our results to locally scrambled unitary ensembles, where we presented explicit formulas for the reconstruction map and the sample complexity bounds. Second, we applied our results to the shadow process tomography of quantum channels with Pauli-invariant unitary ensembles and provided an example where we considered the task of estimating Pauli channels. These general results pave the way for more efficient or robust protocols for predicting pertinent properties of quantum states, including their fidelity, entanglement entropy and quantum fisher information.
ACKNOWLEDGMENTS

K.B. thanks Hong-Ye Hu for the discussion on classical shadows and its connection with quantum entanglement. This work was supported in part by the ARO Grant W911NF-19-1-0302 and the ARO MURI Grant W911NF-20-1-0082.

[1] H.-Y. Huang, R. Kueng, and J. Preskill, Predicting many properties of a quantum system from very few measurements, Nature Physics 16, 1050 (2020).
[2] S. Aaronson, Shadow tomography of quantum states, SIAM Journal on Computing 49, STOC18 (2019).
[3] J. Haah, A. W. Harrow, Z. Ji, X. Wu, and N. Yu, Sample-optimal tomography of quantum states, IEEE Transactions on Information Theory 63, 5628 (2017).
[4] R. O’Donnell and J. Wright, Efficient quantum tomography, in Proceedings of the forty-eighth annual ACM symposium on Theory of Computing (2016) pp. 899–912.
[5] A. Zhao, N. C. Rubin, and A. Miyake, Fermionic partial tomography via classical shadows, Phys. Rev. Lett. 127, 110504 (2021).
[6] H.-Y. Hu and Y.-Z. You, Hamiltonian-driven shadow tomography of quantum states, Phys. Rev. Research 4, 013054 (2022).
[7] H.-Y. Hu, Y. Choi, and Y.-Z. You, Classical shadow tomography with locally scrambled quantum dynamics, arXiv preprint arXiv:2107.04817 (2021).
[8] R. Levy, D. Luo, and B. K. Clark, Classical shadows for quantum process tomography on near-term quantum computers, arXiv preprint arXiv:2110.02965 (2021).
[9] J. Kunjummen, M. C. Tran, D. Carney, and J. M. Taylor, Shadow process tomography of quantum channels, arXiv preprint arXiv:2110.03629 (2021).
[10] H.-Y. Huang, Learning quantum states from their classical shadows, Nat. Rev. Phys. 4, 81 (2022).
[11] C. Hadfield, S. Bravyi, R. Raymond, and A. Mezzacapo, Measurements of quantum Hamiltonians with locally-biased classical shadows, arXiv preprint arXiv:2006.15788 (2020).
[12] C. Hadfield, Adaptive Pauli shadows for energy estimation, arXiv preprint arXiv:2105.12207 (2021).
[13] A. Elben, R. Kueng, H.-Y. R. Huang, R. van Bijnen, C. Kokail, M. Dalmonote, P. Calabrese, B. Kraus, J. Preskill, P. Zoller, and B. Vermerlsch, Mixed-state entanglement from local randomized measurements, Phys. Rev. Lett. 125, 200501 (2020).
[14] A. Neven, J. Carrasco, V. Vitale, C. Kokail, A. Elben, M. Dalmonote, P. Calabrese, P. Zoller, B. Vermerlsch, R. Kueng, et al., Symmetry-resolved entanglement detection using partial transpose moments, npj Quantum Inf. 7, 1 (2021).
[15] A. Rath, C. Branciard, A. Minguzzi, and B. Vermerlsch, Quantum Fisher information from randomized measurements, Phys. Rev. Lett. 127, 260501 (2021).
[16] G. Li, Z. Song, and X. Wang, VSQL: Variational shadow quantum learning for classification, Proceedings of the AAAI Conference on Artificial Intelligence 35, 8357 (2021).
[17] R. J. Garcia, Y. Zhou, and A. Jaffe, Quantum scrambling with classical shadows, Phys. Rev. Research 3, 033155 (2021).
[18] W. J. Huggins, B. A. Gorman, N. C. Rubin, D. R. Reichman, R. Babbush, and J. Lee, Unbiasing fermionic quantum Monte Carlo with a quantum computer, arXiv preprint arXiv:2106.16235 (2021).
[19] H.-Y. Huang, R. Kueng, G. Torlai, V. V. Albert, and J. Preskill, Provably efficient machine learning for quantum many-body problems, arXiv preprint arXiv:2106.12627 (2021).
[20] S. Notarnicola, A. Elben, T. Lahaye, A. Browaeys, S. Montangero, and B. Vermerlsch, A randomized measurement toolbox for Rydberg quantum technologies, arXiv preprint arXiv:2112.11046 (2021).
[21] J. Helsen, M. Ioannou, I. Roth, J. Kitzinger, E. Onorati, A. H. Werner, and J. Eisert, Estimating gate-set properties from random sequences, arXiv preprint arXiv:2110.13178 (2021).
[22] S. H. Sack, R. A. Medina, A. A. Michailidis, R. Kueng, and M. Serbyn, Avoiding barren plateaus using classical shadows, arXiv preprint arXiv:2201.08194 (2022).
[23] S. Chen, W. Yu, P. Zeng, and S. T. Flammia, Robust shadow estimation, PRX Quantum 2, 030348 (2021).
[24] D. E. Koh and S. Grewal, Classical shadows with noise, arXiv preprint arXiv:2011.11580 (2020).
[25] S. T. Flammia, Averaged circuit eigenvalue sampling, arXiv preprint arXiv:2108.05803 (2021).
[26] H.-Y. Huang, R. Kueng, and J. Preskill, Efficient estimation of Pauli observables by derandomization, Phys. Rev. Lett. 127, 030503 (2021).
[27] S. Hillmich, C. Hadfield, R. Raymond, A. Mezzacapo, and R. Wille, Decision diagrams for quantum measurements with shallow circuits, arXiv preprint arXiv:2105.06932 (2021).
[28] B. Wu, J. Sun, Q. Huang, and X. Yuan, Overlapped grouping measurement: A unified framework for measuring quantum states, arXiv preprint arXiv:2105.13091 (2021).
[29] A. Acharya, S. Saha, and A. M. Sengupta, Informationally complete POVM-based shadow tomography, arXiv preprint arXiv:2105.05992 (2021).
[30] J. M. Lukens, K. J. H. Law, and R. S. Bennink, Classical shadows and Bayesian mean estimation: a comparison, in Conference on Lasers and Electro-Optics (Optical Society of America, 2021) p. FW3N.3.
[31] J. M. Lukens, K. J. H. Law, and R. S. Bennink, A Bayesian analysis of classical shadows, npj Quantum Inf. 7, 1 (2021).
[32] S. Chen, J. Cotler, H.-Y. Huang, and J. Li, Exponential separations between learning with and without quantum memory, arXiv preprint arXiv:2111.05881 (2021).
[33] A. Harrow and S. Mehraban, Approximate unitary t-designs by short random quantum circuits using nearest-neighbor and long-range gates, arXiv preprint arXiv:1809.06957 (2018).
[34] Y.-Z. You, Z. Yang, and X.-L. Qi, Machine learning spatial geometry from entanglement features, Phys. Rev. B 97, 045153 (2018).
[35] Y.-Z. You and Y. Gu, Entanglement features of random Hamiltonian dynamics, Phys. Rev. B 98, 014309 (2018).
[36] J. Preskill, Quantum computing in the NISQ era and beyond, Quantum 2, 79 (2018).
[37] K. Bharti, A. Cervera-Lierta, T. H. Kyaw, T. Haug, S. Alperin-Lea, A. Anand, M. Degroote, H. Heimonen, J. S. Kottmann, T. Menke, et al., Noisy intermediate-scale quantum (NISQ) algorithms, arXiv preprint arXiv:2101.08448 (2021).
[38] F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends, R. Biswas, S. Boixo, F. G. Brandao, D. A. Buell, et al., Quantum supremacy using a programmable superconducting processor, Nature 574, 505 (2019).
[39] S. T. Flammia and J. J. Wallman, Efficient estimation of Pauli channels, ACM Transactions on Quantum Computing 1, 1 (2020).
[40] A. Fujiwara and H. Imai, Quantum parameter estimation of a generalized Pauli channel, Journal of Physics A: Mathematical and General 36, 8093 (2003).
[41] M. Hayashi, Quantum channel estimation and asymptotic bound, Journal of Physics: Conference Series 233, 012016 (2010).
[42] A. Chiuri, V. Rosati, G. Vallone, S. Pádua, H. Imai, S. Giacomini, C. Macchiavello, and P. Mataloni, Experimental realization of optimal noise estimation for a general Pauli channel, Phys. Rev. Lett. 107, 253602 (2011).
[43] L. Ruppert, D. Virosztek, and K. Hangos, Optimal parameter estimation of Pauli channels, Journal of Physics A: Mathematical and Theoretical 45, 265305 (2012).
[44] D. Collins, Mixed-state Pauli-channel parameter estimation, Phys. Rev. A 87, 032301 (2013).
[45] R. Harper, W. Yu, and S. T. Flammia, Fast estimation of sparse quantum noise, PRX Quantum 2, 010322 (2021).
[46] S. T. Flammia and R. O’Donnell, Pauli error estimation via Population Recovery, Quantum 5, 549 (2021).
[47] S. Chen, S. Zhou, A. Seif, and L. Jiang, Quantum advantages for Pauli channel estimation, arXiv preprint arXiv:2108.08488 (2020).
[48] A. Roy and A. J. Scott, Unitary designs and codes, Designs, codes and cryptography 53, 13 (2009).
[49] Z. Webb, The Clifford group forms a unitary 3-design, Quantum Inf. Comput. 16, 1379 (2016).
Appendix A: Symplectic Fourier Transform

Let \( V = \mathbb{Z}_2 \times \mathbb{Z}_2 \). We shall denote the \( n \)-fold Cartesian product of \( V \) by \( V^n = V \times V \times \cdots \times V \) and write its elements as \( \vec{x} = (x_1, x_2, \ldots, x_n) \), where each \( x_i = (x_{i1}, x_{i2}) \) is a two-tuple. Given two functions \( f, g : V^n \to \mathbb{R} \), their inner product is defined as

\[
\langle f, g \rangle = \mathbb{E}_{\vec{x} \in V^n} [f(\vec{x})g(\vec{x})].
\] (A1)

where the expectation is taken over the uniform distribution, i.e. \( \mathbb{E}_{\vec{x} \in V^n} = \frac{1}{4^n} \sum_{\vec{x} \in V^n} \).

Next, let us consider the function \( f : V^n \to \mathbb{R} \). Then it is straightforward to show that \( f \) can be written as

\[
f(x) = \sum_{\vec{u} \in V^n} \hat{f}(\vec{u})(-1)^{\langle \vec{u}, \vec{x} \rangle_s},
\] (A2)

where \( \hat{f} : V^n \to \mathbb{R} \) is the symplectic Fourier transform of \( f \) defined as

\[
\hat{f}(\vec{u}) = \mathbb{E}_{\vec{x} \in V^n} [f(\vec{x})(-1)^{\langle \vec{u}, \vec{x} \rangle_s}],
\] (A3)

where the expectation is taken over the uniform distribution, i.e. \( \mathbb{E}_{\vec{x} \in V^n} = \frac{1}{4^n} \sum_{\vec{x} \in V^n} \), and the symplectic inner product between \( \langle \vec{x}, \vec{y} \rangle_s = \sum_i \langle x_i, y_i \rangle_s \) and

\[
\langle x_i, y_i \rangle_s = (a_1, a_2) \cdot (b_1, b_2)_s = a_1 b_2 - a_2 b_1,
\] (A4)

for any \( x_i = (a_1, a_2), y_i = (b_1, b_2) \in V \).

Appendix B: Noiseless classical shadows

**Theorem 14.** [Restatement of Theorem 1] For a Pauli-invariant unitary ensemble \( \mathcal{E} \), the shadow channel can be written as

\[
\mathcal{M}[\rho] = \frac{1}{2^n} \sum_{\vec{d} \in V^n} W_\mathcal{E}[\vec{d}] \text{Tr}[\rho P_{\vec{d}}] P_{\vec{d}},
\] (B1)

where \( W_\mathcal{E}[\vec{d}] \) is the average squared Pauli coefficient of the classical shadow, defined as

\[
W_\mathcal{E}[\vec{d}] = \mathbb{E}_\vec{a} \mathbb{E}_\vec{b} W_{\mathcal{E}, \vec{a}, \vec{b}}[\vec{d}],
\] (B2)

where \( \mathbb{E}_\vec{d} := \frac{1}{4^n} \sum_{\vec{d}} \) denotes the average with respect to the uniform distribution. Hence, for the Pauli-invariant unitary ensemble, the reconstruction map \( \mathcal{M}^{-1} \) exists iff \( W_\mathcal{E}[\vec{d}] > 0 \) for all \( \vec{d} \), and the reconstruction map is defined as follows

\[
\mathcal{M}^{-1}[P_{\vec{d}}] = W_\mathcal{E}[\vec{d}]^{-1} P_{\vec{d}}.
\] (B3)
Proof. Based on the definition of $\mathcal{M}[\rho]$, we have

$$
\mathcal{M}[\rho] = E_U \sum_{\bar{b} \in \{0,1\}^n} \hat{\sigma}_{U,b} \text{Tr}[\hat{\sigma}_{U,b} \rho]
$$

$$
= \sum_b E_U \hat{\sigma}_{U,b} \text{Tr}[\hat{\sigma}_{U,b} \rho]
$$

$$
= \frac{1}{2^n} \sum_{\bar{b} \in \{0,1\}^n} E_U \sum_{\bar{a},\bar{c} \in V^n} \text{Tr}[\hat{\sigma}_{U,b} \rho_a] P_{\bar{a}} \text{Tr}[\hat{\sigma}_{U,b} \rho_{\bar{c}}] P_{\bar{c}} \text{Tr}[P_{\bar{d}} \rho]
$$

$$
= \frac{1}{2^n} \sum_{\bar{b} \in \{0,1\}^n} \sum_{\bar{a},\bar{c} \in V^n} E_U E_{\bar{d} \in V^n} \text{Tr}[\hat{\sigma}_{U,b} \rho_a] P_{\bar{a}} \text{Tr}[\hat{\sigma}_{U,b} \rho_{\bar{c}}] P_{\bar{c}} \text{Tr}[P_{\bar{d}} \rho] \langle -1 \rangle^{(\bar{a},\bar{c},\bar{d})}
$$

$$
= \frac{1}{2^n} \sum_{\bar{a},\bar{c} \in V^n} \sum_{\bar{b} \in \{0,1\}^n} E_U \hat{\sigma}_{U,b} \text{Tr}[\hat{\sigma}_{U,b} \rho_a] P_{\bar{a}} \text{Tr}[\hat{\sigma}_{U,b} \rho_{\bar{c}}] P_{\bar{c}} \text{Tr}[P_{\bar{d}} \rho]
$$

$$
= \frac{1}{2^n} \sum_{\bar{a},\bar{c} \in V^n} \frac{1}{2^n} \sum_{\bar{b} \in \{0,1\}^n} E_U \text{Tr}[\hat{\sigma}_{U,b} \rho_a] P_{\bar{a}} \text{Tr}[\hat{\sigma}_{U,b} \rho_{\bar{c}}] P_{\bar{c}} \text{Tr}[P_{\bar{d}} \rho] \langle -1 \rangle^{(\bar{a},\bar{c},\bar{d})}
$$

$$
= \frac{1}{2^n} \sum_{\bar{a},\bar{c} \in V^n} W_{E}[\bar{a}] \text{Tr}[P_{\bar{d}} \rho] P_{\bar{d}}
$$

(B4)

where the equation (B4) comes from the Pauli decomposition of the classical shadow $\hat{\sigma}_{U,b} = \frac{1}{4} \sum_{\bar{a}} \text{Tr}[\hat{\sigma}_{U,b} \rho_a] P_{\bar{a}}$, the equation (B5) comes from the fact that unitary ensemble is locally Pauli-invariant, the equation (B6) comes from the fact $P_{\bar{a}} P_{\bar{c}} P_{\bar{d}} = (-1)^{(\bar{a},\bar{c},\bar{d})} P_{\bar{a}}$, the equation (B6) comes from the fact $E_{\bar{c}}(-1)^{(\bar{a},\bar{c})} = \delta_{\bar{a},\bar{c}}$. If $W_{E}[\bar{a}] > 0$ for all $\bar{a}$, then the reconstruction map can be defined as follows

$$
\mathcal{M}^{-1}[\rho_{\bar{d}}] = W_{E}[\bar{a}]^{-1} P_{\bar{d}}.
$$

(B9)

Proposition 15 (Restatement of Proposition 2). The coefficients of the reconstruction map can be expressed as follows

$$
W_{E}[S] = (-1)^{|S|} \sum_{A \subset [S]} (-2)^{|A|} E_{E}^{(2)}[A].
$$

(B10)

Proof. It is easy to verify that

$$
W_{E}[S] = (-1)^{|S|} \sum_{A \subset [S]} (-2)^{|A|} E_{E}^{(2)}[A] = (-1)^{|S|} \sum_{A \subset [S]} (-2)^{|A|} \text{Tr}[\hat{\sigma}_{A}^2]
$$

is equivalent to the following equation,

$$
W_{E}[S] = 2^{|S|} \sum_{A \subset S} (-1)^{|A|} \text{Tr}[\hat{\sigma}_{S-A} \hat{\sigma}_{A}^2]
$$

(B11)

where the expectation $E_{\bar{a}}$ denotes the expectation $E_{U,\bar{d}}$, and $\hat{\sigma}_{S}$ denotes the reduced state of $\hat{\sigma}$ on the subset $A$.

For any $n$-qubit state $\rho$, the Pauli coefficient $W_{p}[\bar{a}] = |\text{Tr}[\rho P_{\bar{a}}]|^2$, and

$$
W_{p}[S] = \sum_{\bar{a}, \text{supp}(\bar{a}) = S} |\text{Tr}[\rho P_{\bar{a}}]|^2 = \sum_{\bar{a}, \text{supp}(\bar{a}) = S} |\text{Tr}[\rho_{S} P_{\bar{a}}]|^2,
$$

(B13)

where $\rho_{S}$ is the reduced state on the subset $S$.

Let us take the 1-qubit erasure channel $D[\cdot] = \text{Tr}[\cdot] \mathbb{I}/2$. It is easy to see that $D$ satisfies the following conditions: (1) $D$ is Hermitian, i.e. $D^\dagger = D$; (2) $D$ is idempotent, i.e. $D^2 = D$. For any subset $A \subset [n]$, let us denote $D^A = \otimes_{i \in A} D_i$, where $D_i$ denotes the erasure channel acting on the $i$-th qubit. Now, let us define the 1-qubit linear map $L[\cdot]$ as follows

$$
L[\rho] = \rho - D[\rho].
$$

(B14)
It is easy to see that $L$ satisfies the following conditions (1) $L^1 = L$; (2) $L^2 = L$. For any subset $S \subseteq [n]$, let us denote $L^S = \otimes_{i\in S} L_i$, where $L_j$ denotes the linear map $L$ acting on the $i$-th qubit. Here, we need the following two lemmas—Lemma 16 and 17—which can be verified directly.

**Lemma 16.** The inner product $(L^S[\rho], \rho)$ can be written as
\[
\langle L^S[\rho], \rho \rangle = \frac{1}{2^{|S|}} \sum_{\tilde{a} \in \text{supp}(\tilde{a}) \subseteq S} |\text{Tr}[\rho \tilde{P}_{\tilde{a}}]|^2,
\]
where the inner product between two operators is defined as $\langle A, B \rangle = \text{Tr}[A^\dagger B]$.

**Lemma 17.** The inner product $(L^S[\rho], \rho)$ can be rewritten as
\[
\langle L^S[\rho], \rho \rangle = \sum_{A \subseteq S} (-1)^{|A|} \langle D^A[\rho], \rho \rangle = \sum_{A \subseteq S} (-1)^{|A|} \|D^A[\rho]\|^2_2.
\]

Therefore, based on Lemmas 16 and 17, we have
\[
W_p[S] = 2^{|S|} \sum_{A \subseteq S} (-1)^{|A|} \|D^A[\rho]\|^2_2.
\]

Hence,
\[
W_E[S] = E_d E_{\theta(U,b)} W_{\theta(U,b)}[S] = 2^{|S|} \sum_{A \subseteq S} (-1)^{|A|} E_d E_{\theta(U,b)} \|D^A[\rho]\|^2_2.
\]

**Proposition 18 (Restatement of Proposition 3).** If the observable $O$ is taken to be Pauli operator $P_a$, then the shadow norm is equal to
\[
\|P_a\|^2_{\|E_p\|} = W_E[\tilde{a}]^{-1}.
\]

**Proof.** First,
\[
\|P_a\|^2_{\|E_p\|} = E_p \text{Tr}[M^{-1}[P_a]\hat{\sigma}]^2
\]
\[
= E_U \sum_{b \in \{0,1\}^n} \text{Tr}[M^{-1}[P_a]\hat{\sigma}]^2 \text{Tr}[\hat{\sigma}_{U,b}P]
\]
\[
= E_U \sum_{b \in \{0,1\}^n} E_{\tilde{\sigma}(U,b)} \text{Tr}[P_{\tilde{a}}M^{-1}[P_a]\hat{\sigma}_{U,b}]^2 \text{Tr}[\hat{\sigma}_{U,b}P]
\]
\[
= E_U \sum_{b \in \{0,1\}^n} E_{\tilde{\sigma}(U,b)} \text{Tr}[P_{\tilde{a}}M^{-1}[P_a]\hat{\sigma}_{U,b}]^2 \text{Tr}[\hat{\sigma}_{U,b}P]
\]
\[
= E_U \sum_{b \in \{0,1\}^n} \text{Tr}[M^{-1}[P_a]\hat{\sigma}_{U,b}]^2 \text{Tr}[\hat{\sigma}_{U,b}P]
\]
\[
= E_U \sum_{b \in \{0,1\}^n} \text{Tr}[M^{-1}[P_a]\hat{\sigma}_{U,b}]^2 \text{Tr}[\hat{\sigma}_{U,b}P]
\]
\[
= \frac{1}{2^n} E_U \sum_{b \in \{0,1\}^n} \text{Tr}[M^{-1}[P_a]\hat{\sigma}_{U,b}]^2
\]
\[
= \frac{1}{2^n} E_U \sum_{b \in \{0,1\}^n} W_E[\tilde{a}]^{-2} \text{Tr}[P_{\tilde{a}}\hat{\sigma}_{U,b}]^2
\]
\[
= W_E[\tilde{a}]^{-1},
\]
where the equation (B20) comes from the fact that the unitary ensemble is Pauli-invariant, and the equation (B21) comes from the fact that $P_{\tilde{a}}P_{\tilde{a}}P_{\tilde{a}} = (-1)^{|\tilde{a}|}P_{\tilde{a}}$, the equation (B22) comes from the fact that Pauli group is 1-design\(^1\), the equation (B23) comes the definition of reconstruction map $M^{-1}[P_a] = W_E[\tilde{a}]^{-1}P_{\tilde{a}}$, and the equation (B24) comes the definition of $W_E[\tilde{a}]$.

\(^1\) Note that while the Pauli group is a 1-design, it fails to be 2-design [48, 49].
Proposition 19. [Restatement of Proposition 4] The average shadow norm over the Pauli group $\|O\|^2_Z$ can be expressed as

$$\|O\|^2_Z = \frac{1}{4^n} \sum_{d \in V^n} W_\varepsilon[d]^{-1} W_O[d], \quad (B25)$$

where $W_O[d] = |\text{Tr}[OP_d]|^2$.

Proof. Based on the definition of average shadow norm, we have

$$||O||^2_Z = E_{V \in \mathcal{P}_n} ||O||^2_{V \in \mathcal{P}_n} = E_U \sum_b \text{Tr} \left[ M^{-1} |O| \hat{a} \right]^2 \text{Tr} \left[ \delta_{U,b} \frac{I}{2n} \right]$$

$$= E_U \sum_b \text{Tr} \left[ M^{-1} |O| \hat{a} \right]^2 \text{Tr} \left[ \delta_{U,b} \frac{I}{2n} \right]$$

$$= E_U E_\delta \text{Tr} \left[ M^{-1} |O| \hat{a} \right]^2$$

$$= E_U E_\delta \frac{1}{4^n} \sum_{d, \tilde{d}} \text{Tr} \left[ M^{-1} |O| P_d \right] \text{Tr} [P_\delta \hat{a}] \text{Tr} \left[ M^{-1} |O| P_{\tilde{d}} \right] \text{Tr} [P_\delta \hat{a}]$$

$$= E_U E_\delta \frac{1}{4^n} \sum_{d, \tilde{d}} \text{Tr} \left[ M^{-1} |O| P_d \right] \text{Tr} [P_\delta \hat{a}] \text{Tr} \left[ M^{-1} |O| P_{\tilde{d}} \right] \text{Tr} [P_\delta \hat{a}] \delta_{d, \tilde{d}}$$

$$= E_U E_\delta \frac{1}{4^n} \sum_d \text{Tr} \left[ M^{-1} |O| P_d \right] |^2 \text{Tr} [P_\delta \hat{a}] |^2$$

$$= \frac{1}{4^n} \sum_{d} \left| \text{Tr} \left[ M^{-1} |O| P_d \right] \right|^2 W_\varepsilon[d]$$

$$= \frac{1}{4^n} \sum_{d} W_\varepsilon[d]^{-1} |\text{Tr}[OP_d]|^2$$

$$= \frac{1}{4^n} \sum_{d} W_\varepsilon[d]^{-1} W_O[d].$$

\[\Box\]

Appendix C: Noisy classical shadows

Theorem 20 (Restatement of Theorem 5). Given a Pauli-invariant unitary ensemble $\mathcal{E}$ and a noise channel $\Lambda$, then shadow channel is given by

$$M_{\Lambda}[\rho] = \frac{1}{2^n} \sum_{\bar{a}} W_{\varepsilon_\Lambda}[\bar{a}] \text{Tr} [\rho P_{\bar{a}}] P_{\bar{a}}, \quad (C1)$$

where $W_{\varepsilon_\Lambda}[\bar{a}]$ is the average Pauli coefficient of the noisy classical shadow, and is defined as follows

$$W_{\varepsilon_\Lambda}[\bar{a}] = E_b E_u \text{Tr} \left[ \hat{a} \delta_{U,b} P_{\bar{a}} \right] \text{Tr} \left[ U^\dagger \Lambda U^\dagger |b\rangle \langle b| U P_{\bar{a}} \right]. \quad (C2)$$

Hence, for the Pauli-invariant unitary ensemble, $M^{-1}_{\Lambda}$ exists iff $W_{\varepsilon_\Lambda}[\bar{a}] > 0$ for all $\bar{a}$, and the reconstruction map is defined as follows

$$M^{-1}_{\Lambda}[P_{\bar{a}}] = W_{\varepsilon_\Lambda}[\bar{a}]^{-1} P_{\bar{a}}. \quad (C3)$$
Proof. First,
\[
\mathcal{M}[\rho] = \sum_{\bar{b} \in \{0,1\}^n} \mathbb{E}_{\bar{b}} \sum_{\bar{d}, \bar{c} \in V^n} \mathbb{E}_{\bar{d}} \mathbb{E}_{\bar{c}} \text{Tr} \left[ [\tilde{b}] \bar{b} ] \Lambda [U \rho U^\dagger] \right]
\]
\[
= \frac{1}{2^n} \sum_{\bar{b} \in \{0,1\}^n} \mathbb{E}_{\bar{b}} \sum_{\bar{d}, \bar{c} \in V^n} \text{Tr} \left[ \sigma_{U,\bar{b}} P_{\bar{d}} \right] \text{Tr} \left[ U^\dagger \Lambda^\dagger [\tilde{b}] \bar{b} ] U P_{\bar{c}} \right] \text{Tr} \left[ \rho P_{\bar{c}} \right] \left( -1 \right)^{\langle \bar{d}, \bar{a} \rangle + \langle \bar{c}, \bar{b} \rangle}
\]
\[
= \frac{1}{2^n} \sum_{\bar{b} \in \{0,1\}^n} \mathbb{E}_{\bar{b}} \sum_{\bar{d}, \bar{c} \in V^n} \text{Tr} \left[ \sigma_{U,\bar{b}} P_{\bar{d}} \right] \text{Tr} \left[ U^\dagger \Lambda^\dagger [\tilde{b}] \bar{b} ] U P_{\bar{c}} \right] \text{Tr} \left[ \rho P_{\bar{c}} \right] \delta_{\bar{a},\bar{d}}
\]
\[
= \frac{1}{2^n} \sum_{\bar{b} \in \{0,1\}^n} \mathbb{E}_{\bar{b}} \sum_{\bar{d}, \bar{c} \in V^n} \text{Tr} \left[ \sigma_{U,\bar{b}} P_{\bar{d}} \right] \text{Tr} \left[ U^\dagger \Lambda^\dagger [\tilde{b}] \bar{b} ] U P_{\bar{c}} \right] \text{Tr} \left[ \rho P_{\bar{c}} \right]
\]
\[
= \frac{1}{2^n} \sum_{\bar{a} \in V^n} W_{\lambda}[\bar{a}] \text{Tr} \left[ \rho P_{\bar{a}} \right] P_{\bar{a}},
\]
where \( W_{\lambda}[\bar{a}] \) is defined as
\[
W_{\lambda}[\bar{a}] = \mathbb{E}_{\bar{b}} \mathbb{E}_{\bar{d}} \text{Tr} \left[ \sigma_{U,\bar{b}} P_{\bar{d}} \right] \text{Tr} \left[ U^\dagger \Lambda^\dagger [\tilde{b}] \bar{b} ] U P_{\bar{c}} \right] \text{Tr} \left[ \rho P_{\bar{c}} \right].
\]

(C4)

Proposition 21 (Restatement of Proposition 6). If the observable \( O \) is taken to be Pauli operator \( P_{\bar{a}} \), then the shadow norm is equal to
\[
\| P_{\bar{a}} \|_{\tilde{\lambda}, \rho}^2 = W_{\lambda}[\bar{a}]^{-2} W_{\lambda}[\bar{a}],
\]
where \( W_{\lambda}[\bar{a}] \) is defined as
\[
W_{\lambda}[\bar{a}] = \mathbb{E}_{\bar{b}} \text{Tr} \left[ \sigma_{U,\bar{b}} \right] \text{Tr} \left[ [\tilde{b}] \bar{b} ] U \rho \right].
\]

(C5)

Hence, if \( \Lambda \) is unital, then \( W_{\lambda}[\bar{a}] = W_{\tilde{\lambda}}[\bar{a}] \).

Proof. First,
\[
\| P_{\bar{a}} \|_{\tilde{\lambda}, \rho}^2 = \mathbb{E}_{\bar{b}} \text{Tr} \left[ \mathcal{M}_{\lambda}^{-1} P_{\bar{a}} \sigma \right]^2
\]
\[
= \mathbb{E}_{\bar{b}} \sum_{\bar{b} \in \{0,1\}^n} \text{Tr} \left[ \mathcal{M}_{\lambda}^{-1} P_{\bar{a}} \sigma \right]^2 \text{Tr} \left[ U^\dagger \Lambda^\dagger [\tilde{b}] \bar{b} ] U \rho \right]
\]
\[
= \mathbb{E}_{\bar{b}} \sum_{\bar{b} \in \{0,1\}^n} \mathbb{E}_{\bar{c} \in V^n} \text{Tr} \left[ P_{\bar{c}} \mathcal{M}_{\lambda}^{-1} P_{\bar{a}} \sigma_{U,\bar{b}} \right]^2 \text{Tr} \left[ U^\dagger \Lambda^\dagger [\tilde{b}] \bar{b} ] U P_{\bar{c}} \rho P_{\bar{c}} \right]
\]
\[
= \mathbb{E}_{\bar{b}} \sum_{\bar{b} \in \{0,1\}^n} \mathbb{E}_{\bar{c} \in V^n} \left[ (-1)^{\langle \bar{d}, \bar{c} \rangle} \right] \text{Tr} \left[ \mathcal{M}_{\lambda}^{-1} P_{\bar{a}} \sigma_{U,\bar{b}} \right]^2 \text{Tr} \left[ U^\dagger \Lambda^\dagger [\tilde{b}] \bar{b} ] U P_{\bar{c}} \rho P_{\bar{c}} \right]
\]
\[
= \mathbb{E}_{\bar{b}} \sum_{\bar{b} \in \{0,1\}^n} \mathbb{E}_{\bar{c} \in V^n} \left[ \text{Tr} \left[ \mathcal{M}_{\lambda}^{-1} P_{\bar{a}} \sigma_{U,\bar{b}} \right] \right]^2 \text{Tr} \left[ U^\dagger \Lambda^\dagger [\tilde{b}] \bar{b} ] U P_{\bar{c}} \rho P_{\bar{c}} \right]
\]
\[
= \mathbb{E}_{\bar{b}} \sum_{\bar{b} \in \{0,1\}^n} \mathbb{E}_{\bar{c} \in V^n} \left[ \text{Tr} \left[ \mathcal{M}_{\lambda}^{-1} P_{\bar{a}} \sigma_{U,\bar{b}} \right] \right]^2 \text{Tr} \left[ U^\dagger \Lambda^\dagger [\tilde{b}] \bar{b} ] U P_{\bar{c}} \rho P_{\bar{c}} \right]
\]
\[
= \frac{1}{2^n} \mathbb{E}_{\bar{b}} \sum_{\bar{d}, \bar{c} \in V^n} \mathbb{E}_{\bar{d}} \mathbb{E}_{\bar{c}} \text{Tr} \left[ \sigma_{U,\bar{b}} P_{\bar{d}} \right] \text{Tr} \left[ U^\dagger \Lambda^\dagger [\tilde{b}] \bar{b} ] U P_{\bar{c}} \right] \text{Tr} \left[ \rho P_{\bar{c}} \right]
\]
\[
= \frac{1}{2^n} \mathbb{E}_{\bar{b}} \sum_{\bar{d}, \bar{c} \in V^n} \mathbb{E}_{\bar{d}} \mathbb{E}_{\bar{c}} \text{Tr} \left[ \sigma_{U,\bar{b}} P_{\bar{d}} \right] \text{Tr} \left[ U^\dagger \Lambda^\dagger [\tilde{b}] \bar{b} ] U P_{\bar{c}} \right] \text{Tr} \left[ \rho P_{\bar{c}} \right]
\]
\[
= \frac{1}{2^n} \sum_{\bar{a} \in V^n} W_{\tilde{\lambda}}[\bar{a}] \text{Tr} \left[ P_{\bar{a}} \sigma_{U,\bar{b}} \right] \text{Tr} \left[ [\tilde{b}] \bar{b} ] U \rho \right]
\]
\[
= \frac{1}{2^n} \sum_{\bar{a} \in V^n} W_{\tilde{\lambda}}[\bar{a}] \text{Tr} \left[ P_{\bar{a}} \sigma_{U,\bar{b}} \right] \text{Tr} \left[ [\tilde{b}] \bar{b} ] U \rho \right]
\]
\[
= \frac{1}{2^n} \sum_{\bar{a} \in V^n} W_{\tilde{\lambda}}[\bar{a}] \text{Tr} \left[ P_{\bar{a}} \sigma_{U,\bar{b}} \right] \text{Tr} \left[ [\tilde{b}] \bar{b} ] U \rho \right]
\]
\[
= W_{\lambda}[\bar{a}]^{-2} W_{\lambda}[\bar{a}],
\]

(C6)
where
\[
W^n_{\mathcal{E}}[\bar{a}] = \frac{1}{2^n} \mathbb{E}_U \sum_b |\text{Tr} [P_{\bar{a}} \sigma_{U,b}]|^2 \text{Tr} [\langle b | \Lambda | b \rangle].
\]  
(C12)

**Proposition 22** (Restatement of Proposition 7). *The average shadow norm in the noisy classical shadow protocol with the noise channel \(\Lambda\) can be expressed as follows:
\[
\|O\|_{\mathcal{E}}^2 = \frac{1}{4^n} \sum_{\bar{a}} W_{\mathcal{E}}[\bar{a}]^{-2} W^n_{\mathcal{E}}[\bar{a}] W_{O}[\bar{a}].
\]  
(C13)

**Proof.** First,
\[
\|O\|_{\mathcal{E}}^2 = \mathbb{E}_{\mathcal{E} \in \mathcal{P}_n} \|O\|_{\mathcal{E}^{\text{vphantom{l}}}}^2
= \|O\|_{\mathcal{E}^{\text{vphantom{l}}}}^2
= \mathbb{E}_U \sum_b \text{Tr} \left[ \mathcal{M}_X^{-1}[O] \hat{\sigma} \right]^2 \text{Tr} \left[ U^\dagger \Lambda U \sum_{\bar{a}} W_{\mathcal{E}}[\bar{a}]^{-2} W^n_{\mathcal{E}}[\bar{a}] W_{O}[\bar{a}].
\]

\[
= \mathbb{E}_U \sum_{\bar{a}} \left| \text{Tr} \left[ \mathcal{M}_X^{-1}[O] P_{\bar{a}} \right] \right|^2 \left| \text{Tr} \left[ P_{\bar{a}} \hat{\sigma} \right] \right|^2 \text{Tr} [\langle b | \Lambda | b \rangle]
= \frac{1}{4^n} \sum_{\bar{a}} \text{Tr} \left[ \mathcal{M}_X^{-1}[O] P_{\bar{a}} \right]^2 \left| \text{Tr} \left[ P_{\bar{a}} \hat{\sigma} \right] \right|^2 \text{Tr} [\langle b | \Lambda | b \rangle]
= \frac{1}{4^n} \sum_{\bar{a}} W_{\mathcal{E}}[\bar{a}]^{-2} W^n_{\mathcal{E}}[\bar{a}] \text{Tr} [P_{\bar{a}}] .
\]

\[
\Box
\]

**Appendix D: locally scrambled unitary ensembles**

**Proposition 23** (Restatement of Proposition 8). *Given a locally scrambled unitary ensemble \(\mathcal{E}\), then the shadow channel is
\[
\mathcal{M}[\rho] = \frac{1}{2^n} \sum_{S \subseteq [n]} \tilde{W}_S \sum_{\bar{a} : \text{supp}(\bar{a}) = S} \text{Tr} [\rho P_{\bar{a}}] P_{\bar{a}},
\]

where \(\tilde{W}_S\) is defined as \(\tilde{W}_S = E_{\bar{a} : \text{supp}(\bar{a}) = S} W_{\mathcal{E}}[\bar{a}]\). And the reconstruction map is defined as follows
\[
\mathcal{M}^{-1}[P_{\bar{a}}] = \tilde{W}_S [\text{supp}(\bar{a})]^{-1} P_{\bar{a}}.
\]

**Proof.** Since the locally scrambled unitary ensemble is Pauli-invariant, then according to the above theorem, we have
\[
\mathcal{M}[\rho] = \frac{1}{2^n} \sum_{\bar{a}} W_{\mathcal{E}}[\bar{a}] \text{Tr} [\rho P_{\bar{a}}] P_{\bar{a}},
\]

and
\[
W_{\mathcal{E}}[\bar{a}] = \frac{1}{2^n} \sum_b \mathbb{E}_{U} \left( \text{Tr} \left[ \tilde{\sigma}_{U,b} P_{\bar{a}} \right] \right)^2
= \frac{1}{2^n} \sum_b \mathbb{E}_{U} \left( \text{Tr} \left[ V^n \tilde{\sigma}_{U,b} V^{n \dagger} P_{\bar{a}} \right] \right)^2
= \frac{1}{2^n} \sum_b \mathbb{E}_{\text{supp}(\bar{a})} \left( \text{Tr} \left[ \tilde{\sigma}_{U,b} \otimes P_{\text{supp}(\bar{a})} \right] \right)^2,
\]

\[
\Box
\]
where the expectation $\mathbb{E}_{\text{supp}(\bar{a})}$ is taken over all the Pauli operators with support $\text{supp}(\bar{a})$ without identity component. Hence

$$W_{\mathcal{E}}[\bar{a}] = \mathbb{E}_{\text{supp}(\bar{a})} W_{\mathcal{E}}[\bar{a}] = W_{\mathcal{E}}[\text{supp}(\bar{a})].$$  \hspace{1cm} (D2)

**Proposition 24** (Restatement of Proposition 9). *Given the reconstruction map $\mathcal{M}^{-1}[\sigma] = \sum_s r_s D^s[\sigma]$, then $r_s$ can be expressed by $W_{\mathcal{E}}[A]$ as follows

$$r_s = \sum_{A \subset S} (-1)^{|S|-|A|} W_{\mathcal{E}}[A^c]^{-1},$$

where $A^c$ denotes the complement of $A$ in $[n]$.  \hspace{1cm} (D3)

**Proof.** Since $\mathcal{M}^{-1}[\sigma] = \sum_s r_s D^s[\sigma]$, where $D(\cdot) = \text{Tr} [\cdot] \mathbb{I}/2$ and $D^s = \otimes_{i \in S} D_i$. Hence

$$\mathcal{M}^{-1} [P_{\bar{a}}] = \sum_s r_s D^s [P_{\bar{a}}] = P_{\bar{a}} \sum_s r_s I[\text{supp}(\bar{a}) \subset S],$$

where $I[\cdot]$ is the indicator function. Therefore, we have

$$\sum_s r_s I[A \subset S^c] = W_{\mathcal{E}}[A]^{-1},$$

which is equivalent to

$$\sum_s r_s I[S \subset A] = W_{\mathcal{E}}[A^c]^{-1}.$$  \hspace{1cm} (D5)

Hence, we have

$$r_s = \sum_{A \subset S} (-1)^{|S|-|A|} W_{\mathcal{E}}[A^c]^{-1},$$

which follows from the inclusion-exclusion principle.  \hspace{1cm} $\square$

**Lemma 25** (Inclusion-exclusion principle). *If $g(A) = \sum_{S \subset A} f(S)$, then we have

$$f(A) = \sum_{S \subset A} (-1)^{|A|-|S|} g(S).$$

**Proposition 26** (Restatement of Proposition 11). *Given a locally scrambled unitary ensemble, the average shadow norm is

$$\|O\|_{\mathcal{E}}^2 = \frac{1}{4^n} \sum_{S \subset [n]} \bar{W}_{\mathcal{E}}[S]^{-1} W_O[S],$$

where $W_O[S] = \sum_{\bar{a} : \text{supp}(\bar{a}) = S} W_O[\bar{a}]$.

**Proof.** Since locally scrambled unitary ensembles are Pauli-invariant, we have

$$\|O\|_{\mathcal{E}}^2 = \frac{1}{4^n} \sum_{\bar{a}} W_{\mathcal{E}}[\bar{a}]^{-1} W_O[\bar{a}],$$

and

$$W_{\mathcal{E}}[\bar{a}] = \frac{1}{2^n} \sum_b \mathbb{E}_U \left( \text{Tr} \left[ \hat{\sigma}_{U,b} P_{\bar{a}} \right] \right)^2$$

$$= \frac{1}{2^n} \sum_b \mathbb{E}_U \mathbb{E}_{V^n} \left( \text{Tr} \left[ V^n \hat{\sigma}_{U,b} (V^n)^\dagger P_{\bar{a}} \right] \right)^2$$

$$= \frac{1}{2^n} \sum_b \mathbb{E}_U \mathbb{E}_{\text{supp}(\bar{a})} \left( \text{Tr} \left[ \hat{\sigma}_{U,b} 1_{\text{supp}(\bar{a})^c} \otimes P_{\text{supp}(\bar{a})} \right] \right)^2,$$

where the expectation $\mathbb{E}_{\text{supp}(\bar{a})}$ is taken over all the Pauli operators on support $\text{supp}(\bar{a})$ without identity component. Hence

$$W_{\mathcal{E}}[\bar{a}] = \mathbb{E}_{\text{supp}(\bar{a})} W_{\mathcal{E}}[\bar{a}] = W_{\mathcal{E}}[\text{supp}(\bar{a})].$$  \hspace{1cm} (D11)

$\square$
Appendix E: Classical shadow channel tomography with Pauli-invariant unitary ensembles

**Proposition 27** (Restatement of Proposition 12). Given two Pauli-invariant unitary ensembles $E_i$ and $E_o$, the shadow channel $\mathcal{M}_{i,o}$ is

$$\mathcal{M}_{i,o}[P_{\vec{a}_i} \otimes P_{\vec{a}_o}] = W_{E_i}[P_{\vec{a}_i}]W_{E_o}[P_{\vec{a}_o}]P_{\vec{a}_i} \otimes P_{\vec{a}_o},$$  \hspace{1cm} (E1)

where $W_{E_i}[P_{\vec{a}_i}] = \mathbb{E}_{U_{\vec{a}_i}}[\text{Tr} \left( \hat{\sigma}_i P_{\vec{a}_i} \right) \hat{\sigma}_i P_{\vec{a}_i}]$ and $W_{E_o}[P_{\vec{a}_o}] = \mathbb{E}_{U_{\vec{a}_o}}[\text{Tr} \left( \hat{\sigma}_o P_{\vec{a}_o} \right) \hat{\sigma}_o P_{\vec{a}_o}]$. That is, $\mathcal{M}_{i,o} = \mathcal{M}_i \otimes \mathcal{M}_o$, where $\mathcal{M}_i[P_{\vec{a}_i}] = W_{E_i}[P_{\vec{a}_i}]P_{\vec{a}_i}$ and $\mathcal{M}_o[P_{\vec{a}_o}] = W_{E_o}[P_{\vec{a}_o}]P_{\vec{a}_o}$. Hence, for the Pauli-invariant unitary ensemble, $\mathcal{M}_{i,o}^{-1}$ exists iff $W_{E_i}[\vec{a}_i] > 0, W_{E_o}[\vec{a}_o] > 0$ for all $\vec{a}$, and the reconstruction map is defined as follows

$$\mathcal{M}_{i,o}^{-1}[P_{\vec{a}_i} \otimes P_{\vec{a}_o}] = W_{E_i}^{-1}[\vec{a}_i]^{-1}W_{E_o}^{-1}[\vec{a}_o]^{-1}P_{\vec{a}_i} \otimes P_{\vec{a}_o},$$  \hspace{1cm} (E2)

**Proof.** The proof is similar to the proof of Theorem 14, but we provide it here for the completeness.

\[
\begin{align*}
\mathcal{M}_{i,o}[P_{\vec{a}_i} \otimes P_{\vec{a}_o}] &= \mathbb{E}_{U_{\vec{a}_i}}[\mathbb{E}_{U_{\vec{a}_o}}[\text{Tr} \left( \hat{\sigma}_i P_{\vec{a}_i} \otimes \hat{\sigma}_o P_{\vec{a}_o} \right)]] \\
&= \frac{1}{2^{2n}} \sum_{b_{\vec{a}_i},b_{\vec{a}_o}} \mathbb{E}_{U_{\vec{a}_i}}[\text{Tr} \left( \hat{\sigma}_i P_{\vec{a}_i} \otimes \hat{\sigma}_o P_{\vec{a}_o} \right)] \\
&= \frac{1}{2^{2n}} \sum_{b_{\vec{a}_i},b_{\vec{a}_o}} \mathbb{E}_{U_{\vec{a}_i}}[\text{Tr} \left( \hat{\sigma}_i P_{\vec{a}_i} \otimes \hat{\sigma}_o P_{\vec{a}_o} \right) \hat{\sigma}_i P_{\vec{a}_i} \otimes \hat{\sigma}_o P_{\vec{a}_o} \text{Tr} \left( \hat{\sigma}_i P_{\vec{a}_i} \otimes \hat{\sigma}_o P_{\vec{a}_o} \right)] \\
&= \frac{1}{2^{2n}} \sum_{b_{\vec{a}_i},b_{\vec{a}_o}} \mathbb{E}_{U_{\vec{a}_i}}[\text{Tr} \left( \hat{\sigma}_i P_{\vec{a}_i} \otimes \hat{\sigma}_o P_{\vec{a}_o} \right) \hat{\sigma}_i P_{\vec{a}_i} \otimes \hat{\sigma}_o P_{\vec{a}_o} \text{Tr} \left( \hat{\sigma}_i P_{\vec{a}_i} \otimes \hat{\sigma}_o P_{\vec{a}_o} \right)] (-1)^{(\vec{a}_i,\vec{a}_i) + (\vec{a}_o,\vec{a}_o)} \\
&= \frac{1}{2^{2n}} \sum_{b_{\vec{a}_i},b_{\vec{a}_o}} \mathbb{E}_{U_{\vec{a}_i}}[\text{Tr} \left( \hat{\sigma}_i P_{\vec{a}_i} \otimes \hat{\sigma}_o P_{\vec{a}_o} \right) \hat{\sigma}_i P_{\vec{a}_i} \otimes \hat{\sigma}_o P_{\vec{a}_o} \text{Tr} \left( \hat{\sigma}_i P_{\vec{a}_i} \otimes \hat{\sigma}_o P_{\vec{a}_o} \right)] \\
&= \frac{1}{2^{2n}} \sum_{b_{\vec{a}_i},b_{\vec{a}_o}} \mathbb{E}_{U_{\vec{a}_i}}[\text{Tr} \left( \hat{\sigma}_i P_{\vec{a}_i} \otimes \hat{\sigma}_o P_{\vec{a}_o} \right) \hat{\sigma}_i P_{\vec{a}_i} \otimes \hat{\sigma}_o P_{\vec{a}_o} \text{Tr} \left( \hat{\sigma}_i P_{\vec{a}_i} \otimes \hat{\sigma}_o P_{\vec{a}_o} \right)] \\
&= \frac{1}{2^{2n}} \sum_{b_{\vec{a}_i},b_{\vec{a}_o}} \mathbb{E}_{U_{\vec{a}_i}}[\text{Tr} \left( \hat{\sigma}_i P_{\vec{a}_i} \otimes \hat{\sigma}_o P_{\vec{a}_o} \right) \hat{\sigma}_i P_{\vec{a}_i} \otimes \hat{\sigma}_o P_{\vec{a}_o} \text{Tr} \left( \hat{\sigma}_i P_{\vec{a}_i} \otimes \hat{\sigma}_o P_{\vec{a}_o} \right)] \\
&= \frac{1}{2^{2n}} \sum_{b_{\vec{a}_i},b_{\vec{a}_o}} \mathbb{E}_{U_{\vec{a}_i}}[\text{Tr} \left( \hat{\sigma}_i P_{\vec{a}_i} \otimes \hat{\sigma}_o P_{\vec{a}_o} \right) \hat{\sigma}_i P_{\vec{a}_i} \otimes \hat{\sigma}_o P_{\vec{a}_o} \text{Tr} \left( \hat{\sigma}_i P_{\vec{a}_i} \otimes \hat{\sigma}_o P_{\vec{a}_o} \right)] \\
&= \frac{1}{2^{2n}} \sum_{b_{\vec{a}_i},b_{\vec{a}_o}} \mathbb{E}_{U_{\vec{a}_i}}[\text{Tr} \left( \hat{\sigma}_i P_{\vec{a}_i} \otimes \hat{\sigma}_o P_{\vec{a}_o} \right) \hat{\sigma}_i P_{\vec{a}_i} \otimes \hat{\sigma}_o P_{\vec{a}_o} \text{Tr} \left( \hat{\sigma}_i P_{\vec{a}_i} \otimes \hat{\sigma}_o P_{\vec{a}_o} \right)] \\
&= \frac{1}{2^{2n}} \sum_{b_{\vec{a}_i},b_{\vec{a}_o}} \mathbb{E}_{U_{\vec{a}_i}}[\text{Tr} \left( \hat{\sigma}_i P_{\vec{a}_i} \otimes \hat{\sigma}_o P_{\vec{a}_o} \right) \hat{\sigma}_i P_{\vec{a}_i} \otimes \hat{\sigma}_o P_{\vec{a}_o} \text{Tr} \left( \hat{\sigma}_i P_{\vec{a}_i} \otimes \hat{\sigma}_o P_{\vec{a}_o} \right)] \\
&= \mathbb{E}_{U_{\vec{a}_i}}[\text{Tr} \left( \hat{\sigma}_i P_{\vec{a}_i} \otimes \hat{\sigma}_o P_{\vec{a}_o} \right)] \mathbb{E}_{U_{\vec{a}_o}}[\text{Tr} \left( \hat{\sigma}_o P_{\vec{a}_o} \otimes \hat{\sigma}_i P_{\vec{a}_i} \right)] \\
&= W_{E_i}[\vec{a}_i]W_{E_o}[\vec{a}_o] = \mathcal{M}_{i,o}[P_{\vec{a}_i} \otimes P_{\vec{a}_o}].
\end{align*}
\]

**Proposition 28** (Restatement of Proposition 13). If the observable is taken to be a Pauli operator $P_{\vec{a}}$, then the shadow norm is equal to

$$\| P^T \otimes P_{\vec{a}} \|_{E,\vec{a}} = W_{E_{\vec{a}}}^{-1}[\vec{a}]^{-1}W_{E_{\vec{a}}}^{-1}[\vec{a}]^{-1} = W_{E_{\vec{a}}}^{-1}[\vec{a}]^{-1}W_{E_{\vec{a}}}^{-1}[\vec{a}]^{-1}. \hspace{1cm} (E3)$$
Proof. The proof is similar to the proof of Proposition 19, but we present it here for completeness.

\[
\|\rho \otimes P_{\bar{a}}\|_{\mathcal{E}_t} = \mathbb{E}_{\sigma} \Tr \left( \mathcal{M}_{\sigma, o}^{-1}[\rho^T \otimes P_{\bar{a}}] \hat{\sigma} \right)^2 \\
= \mathbb{E}_{U_j} \mathbb{E}_{U_{\bar{a}}} \sum_{b_i, b_{\bar{a}}} \Tr \left( \mathcal{M}_{\sigma, o}^{-1}[\rho^T \otimes P_{\bar{a}}] \hat{\sigma} \right)^2 \Tr \left[ \hat{\sigma} \mathcal{J}[T] \right] \\
= \mathbb{E}_{U_j} \mathbb{E}_{U_{\bar{a}}} \sum_{b_i, b_{\bar{a}}} \left| \Tr \left( \mathcal{M}_{\sigma, o}^{-1}[\rho^T \otimes P_{\bar{a}}] \hat{\sigma} \right) \right|^2 \Tr \left[ \mathcal{M}_{\sigma, o}^{-1}[P_{\bar{a}}] \hat{\sigma} P_{\bar{a}} \right] \Tr \left[ \hat{\sigma} \mathcal{J}[T] \right] \\
= \mathbb{E}_{U_j} \mathbb{E}_{U_{\bar{a}}} \sum_{b_i, b_{\bar{a}}} \left| \Tr \left( \mathcal{M}_{\sigma, o}^{-1}[\rho^T \otimes P_{\bar{a}}] \hat{\sigma} \right) \right|^2 \Tr \left[ \mathcal{M}_{\sigma, o}^{-1}[P_{\bar{a}}] \hat{\sigma} P_{\bar{a}} \right] \left\| \hat{\sigma} \mathcal{J}[T] \right\|_2^n \\
= \frac{1}{2^n} \mathbb{E}_{U_j} \mathbb{E}_{U_{\bar{a}}} \sum_{b_i, b_{\bar{a}}} \left| \Tr \left( \mathcal{M}_{\sigma, o}^{-1}[\rho^T \otimes P_{\bar{a}}] \hat{\sigma} \right) \right|^2 \\
= W_{\sigma, \bar{a}}[\bar{a}]^{-1} \frac{1}{2^n} \mathbb{E}_{U_j} \sum_{b_i} \left| \Tr \left( \mathcal{M}_{\sigma, o}^{-1}[\rho^T \otimes P_{\bar{a}}] \hat{\sigma} \right) \right|^2 \\
= W_{\sigma, \bar{a}}[\bar{a}]^{-1} \frac{1}{2^n} \mathbb{E}_{U_j} \sum_{b_i} \left| \Tr \left( \mathcal{M}_{\sigma, o}^{-1}[\rho^T \otimes P_{\bar{a}}] \hat{\sigma} \right) \right|^2 \\
= W_{\sigma, \bar{a}}[\bar{a}]^{-1} \frac{1}{2^n} \mathbb{E}_{U_j} \sum_{b_i} \left| \Tr \left( \mathcal{M}_{\sigma, o}^{-1}[\rho^T \otimes P_{\bar{a}}] \hat{\sigma} \right) \right|^2 \\
= W_{\sigma, \bar{a}}[\bar{a}]^{-1} \frac{1}{2^n} \mathbb{E}_{U_j} \sum_{b_i} W_{\sigma, \bar{a}}[\bar{a}]^{-2} \left| \Tr \left( \rho^T P_{\bar{a}} \right) \right|^2 \\
= W_{\sigma, \bar{a}}[\bar{a}]^{-1} \frac{1}{2^n} \sum_{\bar{a}_i} W_{\sigma, \bar{a}}[\bar{a}]^{-1} \left| \Tr \left( \rho^T P_{\bar{a}} \right) \right|^2 .
\]