# Classical Shadows With Noise

Dax Enshan Koh[1,2] and Sabee Grewal[3,2]

[1]Institute of High Performance Computing, Agency for Science, Technology and Research (A*STAR), 1 Fusionopolis Way, #16-16 Connexis, Singapore 138632, Singapore

[2]Zapata Computing, Inc., 100 Federal Street, 20th Floor, Boston, Massachusetts 02110, USA

[3]Department of Computer Science, The University of Texas at Austin, Austin, TX 78712, USA

The classical shadows protocol, recently introduced by Huang, Kueng, and Preskill [Nat. Phys. 16, 1050 (2020)], is a quantum-classical protocol to estimate properties of an unknown quantum state. Unlike full quantum state tomography, the protocol can be implemented on near-term quantum hardware and requires few quantum measurements to make many predictions with a high success probability.

In this paper, we study the effects of noise on the classical shadows protocol. In particular, we consider the scenario in which the quantum circuits involved in the protocol are subject to various known noise channels and derive an analytical upper bound for the sample complexity in terms of a shadow seminorm for both local and global noise. Additionally, by modifying the classical post-processing step of the noiseless protocol, we define a new estimator that remains unbiased in the presence of noise. As applications, we show that our results can be used to prove rigorous sample complexity upper bounds in the cases of depolarizing noise and amplitude damping.

## Contents



Dax Enshan Koh: dax_koh@ihpc.a-star.edu.sg
Sabee Grewal: sabee@cs.utexas.edu

## 1 Introduction

Estimating the expectation values of quantum observables with respect to preparable quantum states is an important subroutine in many NISQ[1]-era quantum algorithms that are of potential practical importance [1,2]. These algorithms, which include variational quantum algorithms [3] like the variational quantum eigensolver (VQE) [4] and the quantum approximate optimization algorithm (QAOA) [5], promise wide-ranging applications in, *inter alia*, quantum chemistry [6], quantum metrology [7], and optimization [8]. However, estimation is often the major bottleneck in many of these applications, where the number of measurements required is often too large for the algorithms to achieve the desired accuracy on useful instances using near-term quantum hardware [9, 10]. Thus, developing efficient estimation protocols that can be implemented on near-term quantum hardware is critical to developing applications for NISQ devices.

In a recent breakthrough, Huang, Kueng, and Preskill introduced the *classical shadows protocol* [11], a protocol for estimating many properties of a quantum state with few quantum measurements. The classical shadows protocol is based on the following idea: instead of recovering a full classical description of a quantum state like in full quantum state tomography [12, 13], the protocol aims to learn only a minimal classical sketch—the classical shadow[2]—of the state, which can then later be used to predict functions of the state (e.g., expectation values of observables).

Classical shadows requires minimal quantum resources, yet can efficiently perform useful estimation tasks, making it amenable for use in the NISQ era. For example, classical shadows can efficiently estimate the energy of local Hamiltonians, verify entanglement, and estimate the fidelity between an unknown quantum state and a known quantum pure state (see [11] for more applications). Additionally, rigorous performance guarantees on the protocol—in the form of upper bounds on the required number of samples in terms of error and confidence parameters—have been proved.

An assumption made in the original work by Huang, Kueng, and Preskill (and in some subsequent works by others) is that the unitary operators involved in the protocol can be executed perfectly. In real-world experiments with actual quantum hardware, however, this assumption will almost never hold due to the effects of noise on the quantum systems involved. Hence, for an accurate description of how the classical shadows protocol will perform in practice, it is important to take into account the effects of noise. The main contribution of this work is theoretically addressing how noise affects classical shadows. We derive rigorous sample complexity upper bounds for the most general noise channel, assuming only that the noise is described by a completely positive and trace-preserving linear superoperator. We also show how our results specialize in specific examples, e.g., when the noise is local, or when the noise is described by a depolarizing channel or an amplitude damping channel.

[1] NISQ—coined by Preskill [1]—stands for *noisy intermediate-scale quantum*.

[2] The term 'shadow' comes from Aaronson's work on *shadow tomography* [14].

## 1.1 Main Ideas

### 1.1.1 Review of Classical Shadows

Classical shadows require the ability to perform computational basis measurements and apply a collection of unitary transformations, called the *unitary ensemble*. The choice of unitary ensemble affects both the time complexity and the number of measurements needed (the sample complexity) for the protocol to succeed with small error. For classical shadows to be time efficient, the unitary ensemble must be efficiently classically simulable on computational basis states (which is why considerable focus is given to the Clifford group in this work and in the original work [11]).

Consider the following random process: sample (with respect to some fixed probability distribution) a unitary transformation $U$ from the unitary ensemble $\mathcal{U}$. Apply $U$ to a quantum state $\rho$, and measure the resulting state $U\rho U^\dagger$ in the computational basis to get outcome state $|b\rangle\langle b|$. Finally, classically simulate $|b\rangle\langle b| \mapsto U^\dagger|b\rangle\langle b|U$. For any unitary ensemble, this process is a quantum channel in expectation, which we call the *noiseless shadow channel*:

$$\mathcal{M} : \rho \mapsto \mathop{\mathbb{E}}_{U\sim\mathcal{U}} \sum_{b\in\{0,1\}^n} \langle b|U\rho U^\dagger|b\rangle U^\dagger|b\rangle\langle b|U, \tag{1}$$

where $\mathbb{E}$ denotes the expectation value. A sufficient condition for the noiseless shadow channel to be invertible is that the unitary ensemble is tomographically complete [11].

**Definition 1.1** (tomographically complete)**.** A unitary ensemble $\mathcal{U}$ is *tomographically complete* if for each pair of quantum states $\sigma \neq \rho$, there exists a $U \in \mathcal{U}$ and $b \in \{0,1\}^n$ such that $\langle b|U\sigma U^\dagger|b\rangle \neq \langle b|U\rho U^\dagger|b\rangle$.

The classical shadows protocol works as follows: run the random process above to produce a classical description of $U^\dagger|b\rangle\langle b|U$ (which involves quantum and classical computation) and then apply the inverse shadow channel $\mathcal{M}^{-1}$ (which involves only classical computation). The output $\hat{\rho} \stackrel{\text{def}}{=} \mathcal{M}^{-1}(U^\dagger|b\rangle\langle b|U)$ is called the *noiseless classical shadow*, which is an unbiased estimator of $\rho$, i.e., $\mathbb{E}[\hat{\rho}] = \rho$. Repeat this process to produce many classical shadows, a classical data set that can be used to estimate linear functions of the unknown state $\rho$.

To estimate a linear function $\mathrm{tr}(O\rho)$, one must classically compute $\mathrm{tr}(O\hat{\rho})$ (an unbiased estimator of $\mathrm{tr}(O\rho)$) for each classical shadow from which a median-of-means estimator is constructed (see Algorithm 1 for details). The power of the median-of-means estimator is captured in the following concentration inequality.

**Fact 1.2** (Jerrum et al. [15])**.** *Let $X$ be a random variable with variance $\sigma^2$. Then $K$ independent sample means of size $N = \frac{34\sigma^2}{\varepsilon^2}$ suffice to construct a median-of-means estimator $\hat{\mu}(N,K)$ that obeys*

$$\mathbf{Pr}[|\hat{\mu}(N,K) - \mathbb{E}[X]| \geq \varepsilon] \leq 2\,\mathrm{e}^{-K/2}, \quad \forall\, \varepsilon > 0. \tag{2}$$

At this point, we have an unbiased estimator of $\mathrm{tr}(O\rho)$ which has nice concentration properties. However, the sample complexity depends on the variance $\mathrm{Var}[\mathrm{tr}(O\hat{\rho})]$ (a function of the input state). To prove *a priori* bounds on the sample complexity of classical shadows (i.e., bounds that do not depend on the input state), the authors introduce the *shadow norm* $\|\cdot\|_{\text{shadow}}$, whose square is always an upper bound on the variance of $\mathrm{tr}(O\hat{\rho})$ (i.e., $\mathrm{Var}[\mathrm{tr}(O\hat{\rho})] \leq \|O\|_{\text{shadow}}^2$). Combining this upper bound with Fact 1.2 yields the main result of [11]:

**Theorem 1.3** (Informal version of Theorem 1 in Huang, Kueng, and Preskill [11]). *Classical shadows of size $N$ suffice to estimate $M$ arbitrary linear functions* $\operatorname{tr}(O_1\rho), \ldots, \operatorname{tr}(O_M\rho)$ *up to additive error $\varepsilon$ given that*

$$N \in O\left(\frac{\log(M)}{\varepsilon^2} \max_{1\le i\le M} \|O_i\|_{\text{shadow}}^2\right).$$

The definition of the shadow norm depends on the unitary ensemble used to create the classical shadows. As examples, Huang, Kueng, and Preskill prove that if the unitary ensemble is the Clifford group, then $\|O\|_{\text{shadow}}^2 \le 3\operatorname{tr}(O^2)$. In this case, the sample complexity is

$$N \in O\left(\frac{\log(M)}{\varepsilon^2} \max_{1\le i\le M} \operatorname{tr}(O_i^2)\right). \tag{3}$$

They also prove that if the unitary ensemble is the $n$-fold tensor product of the single-qubit Clifford group and the observable is a Pauli operator $P = P_1 \otimes \cdots \otimes P_n$, then $\|P\|_{\text{shadow}}^2 = 3^{\mathrm{wt}(P)}$, where $\mathrm{wt}(P) = |\{i : P_i \neq \mathbb{I}\}|$. In this case, the sample complexity is

$$N \in O\left(\frac{\log(M)}{\varepsilon^2} \max_{1\le i\le M} 3^{\mathrm{wt}(P_i)}\right). \tag{4}$$

To prove these bounds, Huang, Kueng, and Preskill use the fact that the Clifford group is a 3-design (Definition 2.5) [16,17]. They first express the shadow norm in terms of expectation values taken over the Clifford group, before replacing these expectations by integrals over the Haar measure by using the fact that the Clifford group forms a 3-design. Roughly speaking, this means that the uniform distribution over the Clifford group can duplicate properties of the probability distribution over the Haar measure for polynomials of degree not more than 3. These integrals have simple closed-form expressions, which can then be shown to be bounded by the expressions found in Eq. (3) and Eq. (4).

### 1.1.2 Our Contributions

In this paper, we consider the scenario in which the unitary operators involved in the classical shadows protocol are subject to noise. Specifically, we assume that an error channel $\mathcal{E}$ acts after the unitary operation is performed in the classical shadows protocol.[3]

The randomized measurement process remains the same as the noiseless protocol with the caveat that the system is subject to noise and continues to describe a quantum channel in expectation that we call the *shadow channel with noise $\mathcal{E}$*:

$$\mathcal{M}_{\mathcal{U},\mathcal{E}}(\rho) = \mathop{\mathbb{E}}_{U\sim\mathcal{U}} \sum_{b\in\{0,1\}^n} \langle b|\mathcal{E}(U\rho U^\dagger)|b\rangle U^\dagger |b\rangle\langle b| U.$$

Note that since the shadow channel depends on noise, so does its inverse $\mathcal{M}_{\mathcal{U},\mathcal{E}}^{-1}$ (assuming it exists[4]). Therefore, while the classical shadows protocol remains similar — produce a

[3]This assumption on the noise model is sometimes referred to as the GTM noise assumption, where GTM stands for *gate-independent*, *time-stationary*, and *Markovian* [18–20]. See Section 6 for remarks on the scope and limitations of this assumption.

[4]We prove sufficient conditions for the invertibility of the shadow channel in the noisy setting (see Claim 4.2).

classical description of $U^\dagger|b\rangle\langle b|U$ and apply the inverse shadow channel — there is a necessary algorithmic change to the protocol when noise is present. Namely, in order for the classical shadow to remain an unbiased estimator, the classical post-processing step (i.e., when the inverse shadow channel is applied) must be modified to account for the noise. (We show this formally in Section 3.)

Noise also affects the sample complexity of the classical shadows shadows protocol. With noise present in the system, we prove the following sample complexity bounds, which generalize the main results of [11]. The key high-level takeaway is that the number of samples increases by only polynomial factors, suggesting that, even with noise, classical shadows can be run efficiently. The bounds below are expressed in terms of the completely dephasing channel $\mathrm{diag} : \mathcal{L}(\mathbb{C}^d) \to \mathcal{L}(\mathbb{C}^d)$, which sends all the non-diagonal entries of its input to zero: $\mathrm{diag}(A) = \sum_{i=1}^d |i\rangle\langle i|A|i\rangle\langle i|$.

**Theorem 1.4** (Informal version of Corollary 4.7)**.** *Classical shadows of size $N$ suffice to estimate $M$ arbitrary linear functions $tr(O_1\rho), \ldots, \mathrm{tr}(O_M\rho)$ to additive error $\varepsilon$ when the quantum circuits used in the protocol are subject to the error channel $\mathcal{E}$ given that*

$$N \in O\left(\frac{2^{2n}\log(M)}{\beta^2\varepsilon^2} \max_{1\le i\le M} \mathrm{tr}(O_i^2)\right),$$

*where $\beta = \mathrm{tr}(\mathcal{E} \circ \mathrm{diag})$.*

Here, $\beta$ is the trace (which we define in Section 2) of the quantum channel $\mathcal{E} \circ \mathrm{diag}$, whose explicit form is given by $\beta = \sum_{i=1}^d \langle i|\mathcal{E}(|i\rangle\langle i|)|i\rangle$; but, roughly speaking, one can think of $\beta$ as the "severity of the noise" on the quantum device. For instance, if we choose $\mathcal{E}$ to be the identity channel (i.e., we model our device as noiseless), then $\beta^2 = 2^{2n}$, which recovers the noiseless sample complexity in Eq. (3). Note that $\beta$ cannot be 0, and so the upper bound in Theorem 1.4 is finite. We discuss this when we prove sufficient conditions for the invertibility of the shadow channel (Section 4.1).

We also generalize the sample complexity bounds when the observables of interest are all Pauli operators.

**Theorem 1.5** (Informal version of Corollary 5.5)**.** *Let $\{P_i\}_{i=1}^M$ be a collection of $M$ Pauli operators. Classical shadows of size $N$ suffice to estimate linear functions $\mathrm{tr}(P_1\rho), \ldots, \mathrm{tr}(P_M\rho)$ to additive error $\varepsilon$ when the quantum circuits used in the protocol are subject to quantum channel $\mathcal{E}^{\otimes n}$ given that*

$$N \in O\left(\frac{\log(M)}{\varepsilon^2} \max_{1\le i\le M} \left(\frac{3}{\beta^2}\right)^{\mathrm{wt}(P_i)}\right),$$

*where $\beta = \mathrm{tr}(\mathcal{E} \circ \mathrm{diag})$ and $\mathrm{wt}(P) = |\{i : P_i \neq \mathbb{I}\}|$.*

It is easy to verify that if we choose $\mathcal{E}$ to be the identity channel, then we recover the noiseless bound (Eq. (4)) proved in [11]. It is important to note that, for this result, we assume that noise can be modelled on the device as a tensor product of single-qubit quantum channels. For simplicity, we have assumed that these single-qubit channels are identical (i.e. each qubit is subject to the same noise model); we note, however, that it is straightforward to generalize this to the case where each single-qubit noise channel is different.

In addition to the sample complexity bounds above, we prove several new results. Among these are:

- Tensor product noise cannot affect nice factorization properties of classical shadows with tensor product structure (Section 3.3).

- We prove a simple sufficient condition for the noisy shadow channel to be invertible (Section 4.1).

- If the unitary ensemble is a 2-design, then the shadow channel is a depolarizing channel (even in the presence of a general quantum channel). (Section 4.1).

- We prove nontrivial generalizations of shadow norm upper bounds presented in [11] (Section 4.3 and Section 5.3). For general noise models, the shadow norm ceases to be a norm. It, however, retains the properties of a seminorm. To this end, we shall refer to the 'generalized shadow norm' as the *shadow seminorm*.

- As applications, we show that our results can be used to prove rigorous sample complexity upper bounds in the cases of depolarizing noise and amplitude damping (Section 4.4 and Section 5.4). We consider these noise models as they are good approximate models for quantum noise occurring in real quantum systems [21].

## 1.2 Related Work

### 1.2.1 Property Estimation and Quantum Tomography

There have been numerous works in the literature that can be cast as algorithms for estimating properties of quantum states. These include the following:

- General algorithms, such as *quantum state tomography*, where the goal is to recover a classical description of an unknown quantum state $\rho$, given copies of $\rho$ [22–26]. Amongst these algorithms are sample-optimal protocols that use an asymptotically optimal number of samples but which require entangled measurements that act simultaneously on all the samples [12, 13], and more experimentally friendly protocols that require only single-sample measurements [27–31].

- *Matrix product state tomography*, where it is assumed that the unknown quantum state is well-approximated by a matrix product state with low bond dimensions [32, 33].

- *Multi-scale entanglement renormalization ansatz (MERA) tomography*, for which a method for reconstructing multi-scale entangled states using a small number of efficiently implementable measurements and fast post-processing was developed [34].

- *Neural network tomography*, which trains a classical deep neural network to represent quantum systems [35, 36].

- *Overlapping quantum tomography*, which uses single-qubit measurements performed in parallel and the theory of perfect hash families to reconstruct $k$-qubit reduced density matrices of an $n$-qubit state with at most $e^{O(k)}\log^2(n)$ rounds of parallel measurements [37].

- *Shadow tomography* [14, 38, 39], where the goal is to estimate $\mathrm{tr}(O_1\rho), \ldots, \mathrm{tr}(O_M\rho)$ to $\pm\varepsilon$ accuracy, given a list of observables $O_1, \ldots, O_M$ and copies $\rho$. Classical shadows can be viewed as an efficient algorithm for shadow tomography in the special case that the shadow norm of the observables is small (e.g., observables with low rank).

### 1.2.2 Solving the Measurement Problem

There have been a number of results which focus on reducing the number of measurements required in near-term quantum algorithms (i.e., solving the so-called "measurement problem"). Recently, several methods have been proposed, for example, Pauli grouping [40, 41], unitary partitioning [42, 43], engineered likelihood functions [44, 45], and deep learning models [35]. See Section 3 of [11] for details on how classical shadows compares with other methods. See also [2,44,46] for more details on the "measurement problem" in near-term quantum algorithms.

### 1.2.3 Classical Shadows

A number of works based on classical shadows have appeared since its introduction in [11]. These include a generalization of classical shadows to the fermionic setting [47–49] and the use of classical shadows to estimate expectation values of molecular Hamiltonians [50] and to detect bipartite entanglement in a many-body mixed state by estimating moments of the partially transposed density matrix [51]. In addition, the first experimental implementation of classical shadows was carried out by Struchalin et al. in a quantum optical experiment with high-dimensional spatial states of photons [52].

During the final stages of preparing version 1 [53] of our manuscript, we became aware of recent independent work by Chen, Yu, Zeng, and Flammia [18], who also study ways to counteract noise in the classical shadows protocol. A key difference between their work and ours is that they do not assume that the noise model is known beforehand. Due to this, their strategy involves first learning the noise as a simple stochastic model before compensating for these errors using robust classical post-processing. In our manuscript, we have assumed that the user has modelled the noise on the device before implementing our protocol. This noise characterization can be carried out using efficient learning methods such as [19, 54].

Subsequent to version 1 [53] of our manuscript, several new extensions and applications of classical shadows have been developed [48, 49, 55–90]. Amongst these are extensions of the classical shadows framework to quantum channels [71, 72] and to more general ensembles, like locally scrambled unitary ensembles [69] and Pauli-invariant unitary ensembles [77]. Additional applications of classical shadows include avoiding barren plateaus in variational quantum algorithms [76], quantifying information scrambling [78], and estimating gate set properties [73].

### 1.2.4 Quantum Error Mitigation

The last few years have seen the invention of several quantum error mitigation techniques [91] to suppress errors in NISQ devices, which are prone to errors but yet are not large enough for quantum error correction [1, 92, 93] to be implemented. Among these techniques are extrapolation methods [94–96] (e.g., Richardson extrapolation and exponential extrapolation), Clifford data regression [97], quantum subspace expansion [98], and probabilistic error cancellation (also known as the quasi-probability method) [95, 99]. Like classical shadows, these techniques involve repeated measurements and classical post-processing to obtain an estimator of the desired result. Unlike these techniques, though, our noisy classical shadows protocol incorporates error mitigation directly into the classical post-processing step, without requiring any additional quantum resources in the measurement process.

Some additional comparisons may be drawn between our noisy classical shadows protocol and the quasi-probability method, first proposed by Temme et al. for special channels [95] and then extended by Endo et al. to practical Markovian noise [99]. The central idea behind the quasi-probability method is that for any (invertible) noise channel, its effects can be reversed by probabilistically implementing its inverse, by using the fact that while the inverse of a quantum channel may not be a quantum channel (and hence cannot be implemented physically by applying unitary operations to quantum states), it may be written as a linear combination of quantum channels (called basis operations). Like the quasi-probability method, our noisy classical shadows protocol reverses the effects of noise by effectively implementing the inverse of the noise channel (as part of implementing the inverse of the shadow channel). Unlike the quasi-probability method where the inverse of the noise channel is applied only probabilistically, our noisy classical shadows protocol applies the inverse of the noise channel deterministically. This is possible since the inverse is applied not as a physical operation on a quantum state, but as a mathematical operation on a classical description of a quantum state.

## 2 Mathematical Preliminaries

Throughout this paper, we denote the set of linear operators on a vector space $V$ by $\mathcal{L}(V)$. The sets of Hermitian operators, unitary operators, and density operators on $\mathbb{C}^d$ are denoted by $\mathbb{H}_d$, $\mathbb{U}_d$, and $\mathbb{D}_d$ respectively. We denote the Haar measure on the $d$-dimensional unitary group by $\eta$.

For a linear operator $A$, the *spectral norm* of $A$ is defined as

$$\|A\|_{\mathrm{sp}} = \max_{x\in\mathbb{C}^n, \|x\|=1} \|Ax\| . \tag{5}$$

When $A$ is Hermitian, the spectral norm may be written as

$$\|A\|_{\mathrm{sp}} = \max_{\sigma\in\mathbb{D}_d} |\mathrm{tr}(\sigma A)| . \tag{6}$$

The set of positive integers is denoted by $\mathbb{Z}^+$. The set of integers from 1 to $d$ is denoted as $[d] = \{1, 2, \ldots, d\}$. The Kronecker delta is denoted by $\delta_{xy}$. We will also use the following generalization of the Kronecker delta:

$$\delta_{x_1 x_2 \ldots x_n} = \begin{cases} 1 & \text{if } x_1 = x_2 = \cdots = x_n \\ 0 & \text{otherwise.} \end{cases} \tag{7}$$

### 2.1 Linear Superoperators and Quantum Channels

We briefly review some properties of linear superoperators and quantum channels that will be used in this paper. For a more comprehensive introduction to quantum channels, see [21, 100].

Let $\mathcal{E} : \mathcal{L}(\mathbb{C}^d) \to \mathcal{L}(\mathbb{C}^d)$ be a linear superoperator. We say that $\mathcal{E}$ is a *quantum channel* if it is both completely positive and trace-preserving. We say that $\mathcal{E}$ is *unital* if the identity operator is a fixed point of $\mathcal{E}$, i.e. $\mathcal{E}(I) = I$. Every linear superoperator $\mathcal{E}$ admits a *Kraus representation*:

$$\mathcal{E}(A) = \sum_a J_a A K_a^\dagger. \tag{8}$$

In the special case when $\mathcal{E}$ is also a quantum channel, $\mathcal{E}$ can be written as

$$\mathcal{E}(A) = \sum_a K_a A K_a^\dagger, \tag{9}$$

where

$$\sum_a K_a^\dagger K_a = I. \tag{10}$$

The vector space of linear operators $\mathcal{L}(\mathbb{C}^d)$ is equipped with the Hilbert-Schmidt inner product $\langle A, B\rangle = \mathrm{tr}(A^\dagger B)$, which is the default inner product on $\mathcal{L}(\mathbb{C}^d)$ that we will use in the rest of the paper. We denote the superoperator adjoint of a superoperator $A$ as $A^*$, and reserve the dagger $()^\dagger$ for the operator adjoint: $()^\dagger = \overline{()}^T$, where $()^T$ and $\overline{()}$ denote the operator transpose and complex conjugate with respect to the computational basis.

Next, we define the quantum channels we consider in this work.

**Definition 2.1** (completely dephasing channel)**.** $A \in \mathcal{L}(\mathcal{C}^d)$. $\mathrm{diag} : \mathcal{L}(\mathbb{C}^d) \to \mathcal{L}(\mathbb{C}^d)$ denotes the *completely dephasing channel*:

$$\mathrm{diag}(A) = \sum_{i=1}^d |i\rangle\langle i|A|i\rangle\langle i|. \tag{11}$$

**Definition 2.2** (depolarizing channel)**.** $A \in \mathcal{L}(\mathbb{C}^d)$. The qudit *depolarizing channel* with depolarizing parameter $f \in \mathbb{R}$ is defined by

$$\mathcal{D}_{n,f}(A) = fA + (1-f)\,\mathrm{tr}(A)\frac{\mathbb{I}}{d}. \tag{12}$$

In the definition above, we have allowed the depolarizing parameter to take any arbitrary value $f \in \mathbb{R}$. We note here however that it is typical to restrict the depolarizing parameter to satisfy $f \in [0, 1]$, especially when $\mathcal{D}_{n,f}$ is viewed as an error channel; for $f$ in this range, one could view $\mathcal{D}_{n,f}$ as a quantum channel that leaves density operators $\rho$ unchanged with probability $f$ and replaces $\rho$ with the maximally mixed state $\mathbb{I}/d$ with probability $1-f$. It is interesting to note, though, that it is not necessary for $f \in [0, 1]$ in order for $\mathcal{D}_{n,f}$ to be a quantum channel (i.e. a completely positive and trace preserving map). While $\mathcal{D}_{n,f}$ is trace-preserving for all $f \in \mathbb{R}$, it is easy to show that $\mathcal{D}_{n,f}$ is completely positive if and only if[5]

$$-\frac{1}{d^2-1} \leq f \leq 1. \tag{13}$$

It then follows that $D_{n,f}$ is a quantum channel if and only if Eq. (13) is satisfied. This property will be used later in the discussion of Claim 4.3.

[5]This follows directly from the fact that the eigenvalues of the Choi matrix $J(D_{n,f})$ corresponding to the depolarizing channel [100] are $\frac{1+f(d^2-1)}{d}$ and $\frac{1-f}{d}$.

**Definition 2.3** (amplitude damping channel)**.** The $n$-qubit *amplitude damping channel* with parameter $p \in [0, 1]$ is defined by

$$\mathrm{AD}_{n,p} = \mathrm{AD}_{1,p}^{\otimes n} \tag{14}$$

where

$$\mathrm{AD}_{1,p} : \begin{pmatrix} \rho_{00} & \rho_{01} \\ \rho_{10} & \rho_{11} \end{pmatrix} \mapsto \begin{pmatrix} \rho_{00} + (1-p)\rho_{11} & \sqrt{p}\rho_{01} \\ \sqrt{p}\rho_{10} & p\rho_{11} \end{pmatrix} \tag{15}$$

is the amplitude damping channel on a single qubit, defined by the Kraus operators

$$K_{\mathrm{AD}0} = \begin{pmatrix} 1 & 0 \\ 0 & \sqrt{p} \end{pmatrix}, \qquad K_{\mathrm{AD}1} = \begin{pmatrix} 0 & \sqrt{1-p} \\ 0 & 0 \end{pmatrix}. \tag{16}$$

The trace of a linear superoperator $\mathcal{E} : \mathcal{L}(\mathbb{C}^d) \to \mathcal{L}(\mathbb{C}^d)$ is

$$\mathrm{tr}(\mathcal{E}) = \sum_{ij} \langle E_{ij}, \mathcal{E}(E_{ij}) \rangle, \tag{17}$$

where $E_{ij} = |i\rangle\langle j|$ and $\langle \cdot, \cdot \rangle$ is the Hilbert-Schmidt inner product. Explicitly,

$$\begin{aligned} \mathrm{tr}(\mathcal{E}) &= \sum_{ij} \mathrm{tr}(E_{ij}^\dagger \mathcal{E}(E_{ij})) \\ &= \sum_{ij} \langle i|\mathcal{E}(|i\rangle\langle j|)|j\rangle. \end{aligned} \tag{18}$$

It is straightforward to check that $\mathrm{tr}(\mathcal{E} \circ \mathrm{diag}) = \sum_i \langle i|\mathcal{E}(|i\rangle\langle i|)|i\rangle$, which is a quantity that appears often throughout this work.

## 2.2 $t$-Fold Twirls and $t$-Designs

$t$-designs are an important concept in quantum information processing with wide-ranging applications ranging from tensor networks [101] and quantum speedup [102–104], to decoupling [105] and quantum state encryption [106]. In this subsection, we shall review the definitions and some properties of $t$-fold twirls and $t$-designs that we will use in this paper. Throughout this subsection, we fix $d \in \mathbb{Z}_{\geq 2}$ to be an integer greater than or equal to 2.

**Definition 2.4** (Twirl)**.** Let $\mathcal{U} \subseteq \mathbb{U}_d$ be a set of unitaries and let $t \in \mathbb{Z}^+$. The $t$-fold *twirl* by $\mathcal{U}$ is the map $\Psi_{\mathcal{U},t} : \mathcal{L}(\mathbb{C}^{d^t}) \to \mathcal{L}(\mathbb{C}^{d^t})$ defined by

$$\Psi_{\mathcal{U},t}(A) = \mathop{\mathbb{E}}_{U \sim \mathcal{U}} U^{\otimes t} A \left(U^\dagger\right)^{\otimes t}. \tag{19}$$

We denote the $t$-fold twirl by the Haar-random unitaries by $T_t^{(d)} : \mathcal{L}(\mathbb{C}^{d^t}) \to \mathcal{L}(\mathbb{C}^{d^t})$, i.e. for all $A \in \mathcal{L}(\mathbb{C}^{d^t})$,

$$T_t^{(d)}(A) = \Psi_{U(\mathbb{C}^d),t}(A) = \int \mathrm{d}\eta(U)\; U^{\otimes t} A \left(U^\dagger\right)^{\otimes t}. \tag{20}$$

When $t=2$, the Haar integral in Eq. (20) may be evaluated as

$$T_2^{(d)}(A) = \frac{1}{d^2-1}\left[\operatorname{tr}(A)\left(I - \frac{W}{d}\right) + \operatorname{tr}(WA)\left(W - \frac{I}{d}\right)\right]. \tag{21}$$

where

$$W = \sum_{i,j=1}^{d} |ij\rangle\langle ji| \tag{22}$$

is the swap operator on $\mathbb{C}^d \otimes \mathbb{C}^d$. For a derivation of Eq. (21), see [100, Eq. (7.179)].

The unitary ensembles considered in Section 4 and Section 5 of this paper are $t$-designs, collections of unitaries that reproduce the $t$-fold twirl by the Haar-random unitaries.

**Definition 2.5** ($t$-design)**.** Let $\mathcal{U} \subseteq \mathbb{U}_d$ be a finite set of unitaries, and let $t \in \mathbb{Z}^+$. We say that $\mathcal{U}$ is a *$t$-design* if $\Psi_{\mathcal{U},t} = T_t^{(d)}$, i.e.

$$\mathop{\mathbb{E}}_{U\sim\mathcal{U}} U^{\otimes t} A \left(U^\dagger\right)^{\otimes t} = \int \mathrm{d}\eta(U)\; U^{\otimes t} A \left(U^\dagger\right)^{\otimes t}. \tag{23}$$

Note that if $\mathcal{U}$ is a $t$-design, then it is also an $s$-design for all $s \le t \in \mathbb{Z}^+$. An important example of a 3-design that fails to be a 4-design is the $n$-qubit Clifford group, denoted by $\mathcal{C}_n$ [16, 17, 107].

We now state a useful identity that we will use later in the paper.

**Lemma 2.6.** *Let $\mathcal{U}$ be a qudit 3-design and $\mathcal{E} : \mathcal{L}(\mathbb{C}^d) \to \mathcal{L}(\mathbb{C}^d)$ be a linear superoperator. Let $A, B, C \in \mathcal{L}(\mathbb{C}^d)$ be linear operators. Then*

$$\begin{aligned}
&\mathop{\mathbb{E}}_{U\sim\mathcal{U}} \sum_{b\in[d]} \langle b|\mathcal{E}(UAU^\dagger)|b\rangle\langle b|UBU^\dagger|b\rangle\langle b|UCU^\dagger|b\rangle \\
&= \frac{(1+d)\alpha - 2\beta}{(d-1)d(d+1)(d+2)}\left(\operatorname{tr}(A)\operatorname{tr}(BC) + \operatorname{tr}(A)\operatorname{tr}(B)\operatorname{tr}(C)\right) \\
&\quad + \frac{d\beta - \alpha}{(d-1)d(d+1)(d+2)}\left(\operatorname{tr}(AB)\operatorname{tr}(C) + \operatorname{tr}(AC)\operatorname{tr}(B) + \operatorname{tr}(ABC) + \operatorname{tr}(ACB)\right),
\end{aligned} \tag{24}$$

*where $\alpha = \operatorname{tr}(\mathcal{E}(I))$ and $\beta = \operatorname{tr}(\mathcal{E} \circ \operatorname{diag})$.*

We present a proof of Lemma 2.6 in Appendix A.1.

# 3 Noisy Classical Shadows

## 3.1 Generating the Classical Shadow

We study the setting where an error channel $\mathcal{E}$ acts on the input state right after some $U$ from the unitary ensemble $\mathcal{U}$ is applied. We assume access to a noisy measurement primitive, slightly altered from [11].

**Definition 3.1** (noisy measurement primitive)**.** We can apply a restricted set of unitary transformations $\rho \mapsto U\rho U^\dagger$, where $U$ is chosen uniformly at random from a unitary ensemble $\mathcal{U}$. Subsequently, an error channel $\mathcal{E}$ acts on the state $U\rho U^\dagger \mapsto \mathcal{E}(U\rho U^\dagger)$. Finally, we can measure the state in the computational basis $\{|b\rangle : b \in \{0,1\}^n\}$.

The randomized measurement procedure remains the same as [11] with the caveat that the transformed state is subject to an error channel. The randomized measurement procedure is as follows. Given copies of an input state $\rho$, perform the following on each copy: transform $\rho \mapsto \mathcal{E}(U\rho U^\dagger)$, measure in the computational basis, and apply $U^\dagger$ to the post-measurement state. The output of this procedure is

$$U^\dagger|\hat{b}\rangle\langle\hat{b}|U \quad \text{with probability} \quad P_b(\hat{b}) \overset{\text{def}}{=} \langle\hat{b}|\mathcal{E}(U\rho U^\dagger)|\hat{b}\rangle \quad \text{where} \quad \hat{b} \in \{0,1\}^n. \tag{25}$$

In expectation, this procedure describes a quantum channel.

**Definition 3.2** (shadow channel)**.** Let $\mathcal{M}_{\mathcal{U},\mathcal{E}} : \mathcal{L}(\mathbb{C}^{2^n}) \to \mathcal{L}(\mathbb{C}^{2^n})$ be defined by

$$\mathcal{M}_{\mathcal{U},\mathcal{E}}(\rho) \overset{\text{def}}{=} \mathop{\mathbb{E}}_{U\sim\mathcal{U}} \sum_{b\in\{0,1\}^n} \langle b|\mathcal{E}(U\rho U^\dagger)|b\rangle U^\dagger|b\rangle\langle b|U = \mathop{\mathbb{E}}_{\substack{U\sim\mathcal{U}\\ \hat{b}\sim P_b}} U^\dagger|\hat{b}\rangle\langle\hat{b}|U. \tag{26}$$

We call $\mathcal{M}_{\mathcal{U},\mathcal{E}}$ the shadow channel with noise $\mathcal{E}$.

One can view $\mathcal{M}_{\mathcal{U},\mathcal{E}}$ as the expected output of the random measurement procedure described above. Note that when we take the channel $\mathcal{E}$ to be the identity channel, we recover the *noiseless shadow channel* $\mathcal{M}$ given by Eq. (1).

**Claim 3.3.** *For all unitary ensembles $\mathcal{U}$ and quantum channels $\mathcal{E}$, $\mathcal{M}_{\mathcal{U},\mathcal{E}}$ is a quantum channel.*

*Proof.* $\mathcal{M}_{\mathcal{U},\mathcal{E}} = \mathcal{D}\circ\mathcal{C}\circ\mathcal{B}\circ\mathcal{A}$ is a composition of quantum channels, where $\mathcal{A}(X) = \mathbb{E}_{U\sim\mathcal{U}} UXU^\dagger$ (mixed unitary channel), $\mathcal{B}(X) = \mathcal{E}(X)$ (error channel), $\mathcal{C}(X) = \sum_b |b\rangle\langle b|X|b\rangle\langle b|$ (quantum-to-classical channel), and $\mathcal{D}(X) = U^\dagger XU$ (unitary channel). □

In the noiseless case, it is known that if the unitary ensemble is tomographically complete, then the shadow channel is invertible [11]. Similarly, we prove sufficient conditions for the shadow channel to be invertible in the noisy case (see Claim 4.2). Assuming the shadow channel is invertible, we can define the classical shadow.

**Definition 3.4** (classical shadow)**.** Assuming the shadow channel is invertible with inverse $\mathcal{M}^{-1}_{\mathcal{U},\mathcal{E}}$, define the *classical shadow* $\hat{\rho}$ as

$$\hat{\rho} = \hat{\rho}(\mathcal{U},\mathcal{E},\hat{U},\hat{b}) \overset{\text{def}}{=} \mathcal{M}^{-1}_{\mathcal{U},\mathcal{E}}(\hat{U}^\dagger|\hat{b}\rangle\langle\hat{b}|\hat{U}). \tag{27}$$

The classical shadow is a random matrix with unit trace[6] and reproduces $\rho$ in expectation: $\mathbb{E}[\hat{\rho}] = \rho$.[7] Repeating this process $N$ times produces a classical shadow with size $N$.

[6] $\mathrm{tr}(\hat{\rho}) = \mathrm{tr}(\mathcal{M}^{-1}_{\mathcal{U},\mathcal{E}}(\hat{U}^\dagger|\hat{b}\rangle\langle\hat{b}|\hat{U})) = \mathrm{tr}(\hat{U}^\dagger|\hat{b}\rangle\langle\hat{b}|\hat{U}) = 1$. We use the fact that the inverse shadow channel is trace preserving because the shadow channel is trace preserving.

[7] $\mathbb{E}[\hat{\rho}] = \mathbb{E}[\mathcal{M}^{-1}_{\mathcal{U},\mathcal{E}}(\hat{U}^\dagger|\hat{b}\rangle\langle\hat{b}|\hat{U})] = \mathcal{M}^{-1}_{\mathcal{U},\mathcal{E}}(\mathbb{E}[\hat{U}^\dagger|\hat{b}\rangle\langle\hat{b}|\hat{U}]) = \mathcal{M}^{-1}_{\mathcal{U},\mathcal{E}}(\mathcal{M}_{\mathcal{U},\mathcal{E}}(\rho)) = \rho$.

**Definition 3.5** (size-$N$ classical shadow)**.** The size-$N$ *classical shadow* corresponding to pairs $(U_1, \hat{b}_1), \ldots, (U_N, \hat{b}_N)$ is

$$\mathsf{S}(\rho; N) = \{\hat{\rho}_1, \ldots, \hat{\rho}_N\} \quad \text{where} \quad \hat{\rho}_i = \mathcal{M}_{\mathcal{U},\mathcal{E}}^{-1}(U_i^\dagger |\hat{b}_i\rangle\langle\hat{b}_i| U_i). \tag{28}$$

## 3.2 Noisy Classical Shadows Protocol

The classical shadow is a classical dataset that can be used to predict many linear functions in the unknown state $\rho$. Recall that a linear function in a quantum state $\rho$ is a function of the form $\rho \mapsto \mathrm{tr}(O\rho)$ for some linear operator $O$. It is easy to confirm that the random variable $\mathrm{tr}(O\hat{\rho})$ reproduces $\mathrm{tr}(O\rho)$ in expectation: $\mathbb{E}[\mathrm{tr}(O\hat{\rho})] = \mathrm{tr}(O\,\mathbb{E}[\hat{\rho}]) = \mathrm{tr}(O\rho)$. Therefore, we can use the classical shadow to produce unbiased estimates of $\mathrm{tr}(O_1\rho), \ldots, \mathrm{tr}(O_M\rho)$ for any observables $O_1, \ldots, O_M$. We continue to use median-of-means estimation, as was done in the noiseless protocol:

**Algorithm 1** Median-of-means estimation based on a classical shadow.

**Input:** a list of observables $O_1, \ldots, O_M$, size-$L$ classical shadow $\mathsf{S}(\rho; L)$, $K \in \mathbb{Z}^+$.

1: Set $\hat{\rho}_{(k)} = \frac{1}{\lfloor L/K \rfloor} \sum_{i=(k-1)\lfloor L/K \rfloor+1}^{k\lfloor L/K \rfloor} \hat{\rho}_i$, for $k = 1, \ldots, K$

2: Output $\hat{o}_i(\lfloor L/K \rfloor, K) = \mathrm{median}\left\{ \mathrm{tr}(O_i\hat{\rho}_{(1)}), \ldots, \mathrm{tr}(O_i\hat{\rho}_{(K)}) \right\}$, for $i = 1, \ldots, M$

Using the concentration properties of median-of-means estimators (see Fact 1.2), we understand how estimates $\mathrm{tr}(O\hat{\rho})$ concentrate around the true value as a function of $\mathrm{Var}[\mathrm{tr}(O\hat{\rho})]$. However, we are interested in bounds that are independent of the input state $\rho$, which motivates the following lemma.

**Lemma 3.6.** *Let $\mathcal{U}$ be a set of $n$-qubit unitary transformations and let $\mathcal{E}$ be an $n$-qubit quantum channel. Assume that the shadow channel $\mathcal{M}_{\mathcal{U},\mathcal{E}}$ (Eq. (26)) is invertible. Let $O \in \mathbb{H}_{2^n}$ and $\rho \in \mathbb{D}^{2^n}$ be an unknown $n$-qubit state. Let $\hat{o} = \mathrm{tr}(O\hat{\rho})$, where $\hat{\rho}$ is the classical shadow (Eq. (27)). Then,*

$$\underset{\substack{U\sim\mathcal{U}\\ b\sim P_b}}{\mathrm{Var}}[\hat{o}] \leq \left\| O - \mathrm{tr}(O)\frac{\mathbb{I}}{2^n} \right\|_{\mathrm{shadow},\mathcal{U},\mathcal{E}}^2, \tag{29}$$

*where*

$$\|O\|_{\mathrm{shadow},\mathcal{U},\mathcal{E}} = \max_{\sigma\in\mathbb{D}_{2^n}} \sqrt{\underset{U\sim\mathcal{U}}{\mathbb{E}} \sum_{b\in\{0,1\}^n} \langle b|\mathcal{E}(U\sigma U^\dagger)|b\rangle\langle b|U\mathcal{M}_{\mathcal{U},\mathcal{E}}^{-1,\dagger}(O)U^\dagger|b\rangle^2}. \tag{30}$$

We call the function $\|\cdot\|_{\mathrm{shadow},\mathcal{U},\mathcal{E}}$, which depends on only the unitary ensemble and the error channel, the *shadow seminorm*. As we show in Appendix B, the shadow seminorm is indeed a *seminorm*, i.e., it satisfies absolute homogeneity and the triangle inequality. However, unlike the noiseless case [11], the noisy shadow seminorm is not necessarily a norm: there exist noise channels $\mathcal{E}$ for which $\|\cdot\|_{\mathrm{shadow},\mathcal{U},\mathcal{E}}$ fails to satisfy the point-separating property required of a norm. In Appendix B, we also explore the question about when the shadow seminorm is a norm. In particular, we prove that a sufficient condition for it to be a norm is that $\mathcal{E}$

satisfies the following property: for all $b \in \{0,1\}^n$, there exists a density operator $\sigma \in \mathbb{D}(\mathbb{C}^{2^n})$ such that $\langle b|\mathcal{E}(\sigma)|b\rangle \neq 0$.

Again, the motivation for introducing the shadow seminorm is to get an upper bound on $\mathrm{Var}[\hat{o}]$ that does not depend on the unknown state $\rho$. The proof is a straightforward generalization of Lemma S1 in [11], which we defer to Appendix A.2.

Thus far, we have shown that the noisy classical shadow is an unbiased estimator of linear functions in $\rho$ and have proved an upper bound on the variance of the estimator. This is enough to prove the following performance guarantee on the noisy classical shadows protocol.

**Theorem 3.7.** *Fix an $n$-qubit unitary ensemble $\mathcal{U}$, a collection of $n$-qubit observables $O_1, \ldots, O_M$, an $n$-qubit quantum channel $\mathcal{E}$, and accuracy parameters $\varepsilon, \delta \in [0,1]$. Assume that $\mathcal{M}_{\mathcal{U},\mathcal{E}}$ (Eq. (26)) is invertible. Set*

$$K = 2\log\frac{2M}{\delta} \qquad \text{and} \qquad N = \frac{34}{\varepsilon^2}\max_{1\le i\le M}\left\|O_i - \frac{1}{2^n}\operatorname{tr}(O_i)\mathbb{1}\right\|^2_{\mathrm{shadow},\mathcal{U},\mathcal{E}}.$$

*Then, a size-$(NK)$ classical shadow $\mathsf{S}(\rho; NK)$ is sufficient to estimate $\hat{o}_1, \ldots, \hat{o}_M$ with the following performance guarantee:*

$$\mathbf{Pr}\Big[|\hat{o}_i(N,K) - \operatorname{tr}(O_i\rho)| \le \varepsilon \quad \forall i = 1,\ldots,M\Big] \ge 1-\delta. \tag{31}$$

*Hence, the sample complexity to estimate a collection of $M$ linear target functions $\operatorname{tr}(O_i\rho)$ within error $\varepsilon$ and failure probability $\delta$ is*

$$NK = O\left(\frac{\log(M/\delta)}{\varepsilon^2}\max_{1\le i\le M}\left\|O_i - \frac{1}{2^n}\operatorname{tr}(O_i)\mathbb{1}\right\|^2_{\mathrm{shadow},\mathcal{U},\mathcal{E}}\right).$$

*Proof.* Run Algorithm 1 with $O_1, \ldots, O_M$, $\mathsf{S}(\rho; NK)$, $N$, and $K$ to obtain estimates

$$\hat{o}_i(N,K) = \operatorname{median}\left\{\operatorname{tr}(O_i\hat{\rho}_{(1)}, \ldots, \operatorname{tr}(O_i\hat{\rho}_{(K)})\right\}, \qquad \text{for } i = 1,\ldots,M. \tag{32}$$

Then,

$$\begin{aligned}
\mathbf{Pr}\Big[|\hat{o}_i(N,K) - \operatorname{tr}(O_i\rho)| \le \varepsilon\ \forall i = 1,\ldots,M\Big] &= 1 - \mathbf{Pr}\left[\exists i = 1,\ldots,M : |\hat{o}_i(N,K) - \operatorname{tr}(O_i\rho)| \le \varepsilon\right] \\
&\ge 1 - \sum_{i=1}^{M}\mathbf{Pr}\left[|\hat{o}_i(N,K) - \operatorname{tr}(O_i\rho)| > \varepsilon\right] \\
&\ge 1 - 2\,\mathrm{e}^{-K/2}\sum_{i=1}^{M} 1 \\
&= 1-\delta.
\end{aligned} \tag{33}$$

The inequality on the second line follows from the union bound and the inequality on the third line follows from Fact 1.2. $\square$

The noisy classical shadows protocol is summarized next.

Summary: Classical Shadows With Noise

**Hyperparameters**

Let $\mathcal{U}$ be a set of $n$-qubit unitary transformations. Let $\mathcal{E}$ be an $n$-qubit quantum channel.

**Definitions**

Let

$$\mathcal{M}_{\mathcal{U},\mathcal{E}} : \rho \mapsto \mathop{\mathbb{E}}_{U\sim\mathcal{U}} \sum_{b\in\{0,1\}^n} \langle b|\mathcal{E}(U\rho U^\dagger)|b\rangle U^\dagger|b\rangle\langle b|U,$$

and assume that $\mathcal{M}_{\mathcal{U},\mathcal{E}}$ is invertible. Let

$$\|O\|_{\mathrm{shadow},\mathcal{U},\mathcal{E}} = \max_{\sigma\in\mathbb{D}_{2^n}} \sqrt{\mathop{\mathbb{E}}_{U\sim\mathcal{U}} \sum_{b\in\{0,1\}^n} \langle b|\mathcal{E}(U\sigma U^\dagger)|b\rangle\langle b|U\mathcal{M}_{\mathcal{U},\mathcal{E}}^{-1,\dagger}(O)U^\dagger|b\rangle^2}.$$

**Algorithm 2** Classical Shadows With Noise

**Input**

$\rho \in \mathbb{C}^{2^n}$ (an unknown $n$-qubit state, given as multiple copies of a black box). $\varepsilon, \delta \in (0,1)$ (accuracy parameters). $O_1, \ldots, O_M$ (a list of observables).

**Output**

Estimators $\hat{o}_1, \ldots, \hat{o}_M$ such that

$$\mathbf{Pr}\Big[|\hat{o}_i - \mathrm{tr}(O_i\rho)| \le \varepsilon \quad \forall i = 1,\ldots,M\Big] \ge 1-\delta.$$

**Initialization**

Set $K = 2\log\frac{2M}{\delta}$

Set $N = \frac{34}{\varepsilon^2}\max_{1\le i\le M}\left\|O_i - \frac{1}{2^n}\mathrm{tr}(O_i)\mathbb{1}\right\|^2_{\mathrm{shadow},\mathcal{U},\mathcal{E}}$

**Classical shadow generation**

**for** $i = 1,\ldots,NK$ **do**

  Randomly choose $\hat{U}_i \in \mathcal{U}$

  Apply $\rho \mapsto \mathcal{E}(U\rho U^\dagger)$ to (a fresh copy of) $\rho$ to get $\rho_1$

  Perform a computational basis measurement on $\rho_1$ to get outcome $\hat{b}_i \in \{0,1\}^n$

  Save a classical description of $\hat{U}_i^\dagger|\hat{b}_i\rangle\langle\hat{b}_i|\hat{U}_i$ in classical memory

  Apply $\mathcal{M}_{\mathcal{U},\mathcal{E}}^{-1}$ to $\hat{U}_i^\dagger|\hat{b}_i\rangle\langle\hat{b}_i|\hat{U}_i$ to get $\hat{\rho}_i = \mathcal{M}_{\mathcal{U},\mathcal{E}}^{-1}(\hat{U}_i^\dagger|\hat{b}_i\rangle\langle\hat{b}_i|\hat{U}_i)$

Set $\mathsf{S}(\rho; NK) = \{\hat{\rho}_1,\ldots,\hat{\rho}_{NK}\}$

**Median-of-means estimation**

Set $\hat{\rho}_{(k)} = \frac{1}{N}\sum_{i=(k-1)N+1}^{kN}\hat{\rho}_i$, for $k = 1,\ldots,K$

Output $\hat{o}_i \stackrel{\mathrm{def}}{=} \hat{o}_i(N,K) = \mathrm{median}\left\{\mathrm{tr}(O_i\hat{\rho}_{(1)}),\ldots,\mathrm{tr}(O_i\hat{\rho}_{(K)})\right\}$, for $i = 1,\ldots,M$

### 3.3 Product Ensembles with Product Noise

We conclude this section by showing some nice factorization properties for classical shadows when the unitary ensemble is a product ensemble and the quantum channel is a product channel.

**Definition 3.8** (product channel)**.** An $n$-qubit *product channel* is a quantum channel $\mathcal{E}$ of the form $\mathcal{E} = \mathcal{E}_1 \otimes \ldots \otimes \mathcal{E}_n : \mathcal{L}(\mathbb{C}^{2^n}) \to \mathcal{L}(\mathbb{C}^{2^n})$, where each $\mathcal{E}_i : \mathcal{L}(\mathbb{C}^2) \to \mathcal{L}(\mathbb{C}^2)$.

**Definition 3.9** (product ensemble)**.** An $n$-qubit *product ensemble* is a collection of unitary transformations of the form $\mathcal{U} = \bigotimes_{i=1}^{n} \mathcal{U}_i = \{U_1 \otimes \ldots \otimes U_n : U_1 \in \mathcal{U}_1, \ldots, U_n \in \mathcal{U}_n\}$.

First, we show that the shadow channel, the inverse shadow channel, and the classical shadow all factorize into tensor products when the unitary ensemble is a product ensemble and the quantum channel is a product channel.

**Claim 3.10.** *Let $\mathcal{U} = \bigotimes_{i=1}^{n} \mathcal{U}_i$ be a product ensemble, and let $\mathcal{E} = \mathcal{E}_1 \otimes \ldots \otimes \mathcal{E}_n$ be a product channel. Then, the shadow channel factorizes as follows:*

$$\mathcal{M}_{\mathcal{U},\mathcal{E}} = \bigotimes_{i=1}^{n} \mathcal{M}_{\mathcal{U}_i,\mathcal{E}_i}.$$

*Assume that $\mathcal{M}_{\mathcal{U}_i,\mathcal{E}_i}$ is invertible $\forall i \in \{1, \ldots, n\}$. Then, $\mathcal{M}_{\mathcal{U},\mathcal{E}}$ is invertible and the inverse shadow channel factorizes as follows:*

$$\mathcal{M}^{-1}_{\mathcal{U},\mathcal{E}} = \bigotimes_{i=1}^{n} \mathcal{M}^{-1}_{\mathcal{U}_i,\mathcal{E}_i}.$$

*Finally, let $\hat{U} = \hat{U}_1 \otimes \ldots \otimes \hat{U}_n \in \mathcal{U}$ and $\hat{b} = \hat{b}_1 \ldots \hat{b}_n \in \{0,1\}^n$. Then,*

$$\hat{\rho}(\mathcal{U}, \mathcal{E}, \hat{U}, \hat{b}) = \bigotimes_{i=1}^{n} \hat{\rho}(\mathcal{U}_i, \mathcal{E}_i, \hat{U}_i, \hat{b}_1). \tag{34}$$

*Proof.* The first part of the claim follows from two basic facts: Tensor products of quantum channels factorize when applied to elementary tensor products, and the $n$-th order tensor product is the linear hull of all elementary tensor products. The second part of the claim follows from the fact that $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$.

The third part of the claim follows from the following chain of equalities:

$$\hat{\rho}(\mathcal{U}, \mathcal{E}, \hat{U}, \hat{b}) = \mathcal{M}^{-1}_{\mathcal{U},\mathcal{E}}(\hat{U}^\dagger |\hat{b}\rangle\langle\hat{b}| \hat{U}) = \bigotimes_{i=1}^{n} \mathcal{M}^{-1}_{\mathcal{U}_i,\mathcal{E}_i}(\hat{U}_i^\dagger |\hat{b}_i\rangle\langle\hat{b}_i| \hat{U}_i) = \bigotimes_{i=1}^{n} \hat{\rho}(\mathcal{U}_i, \mathcal{E}_i, \hat{U}_i, \hat{b}_1). \tag{35}$$

□

**Claim 3.11.** *Let $\mathcal{U} = \bigotimes_{i=1}^{n} \mathcal{U}_i$ be a product ensemble, and let $\mathcal{E} = \mathcal{E}_1 \otimes \ldots \otimes \mathcal{E}_n$ be a product channel. Assume that $\mathcal{M}_{\mathcal{U}_i,\mathcal{E}_i}$ is invertible $\forall i \in \{1, \ldots, n\}$. Let $\hat{U} = \hat{U}_1 \otimes \ldots \otimes \hat{U}_n \in \mathcal{U}$ and $\hat{b} = \hat{b}_1 \ldots \hat{b}_n \in \{0,1\}^n$. Then,*

$$\hat{\rho}(\mathcal{U}, \mathcal{E}, \hat{U}, \hat{b}) = \bigotimes_{i=1}^{n} \hat{\rho}(\mathcal{U}_i, \mathcal{E}_i, \hat{U}_i, \hat{b}_1). \tag{36}$$

*Proof.* By Claim 3.10, given a product ensemble and product channel, the shadow channel is $\mathcal{M}_{\mathcal{U},\mathcal{E}} = \bigotimes_{i=1}^n \mathcal{M}_{\mathcal{U}_i,\mathcal{E}_i}$, and $\mathcal{M}_{\mathcal{U},\mathcal{E}}^{-1} = \bigotimes_{i=1}^n \mathcal{M}_{\mathcal{U}_i,\mathcal{E}_i}^{-1}$.

$$\begin{aligned}\hat{\rho}(\mathcal{U},\mathcal{E},\hat{U},\hat{b}) &= \mathcal{M}_{\mathcal{U},\mathcal{E}}^{-1}(\hat{U}^\dagger|\hat{b}\rangle\langle\hat{b}|\hat{U}) \\ &= \bigotimes_{i=1}^n \mathcal{M}_{\mathcal{U}_i,\mathcal{E}_i}^{-1}(\hat{U}_i^\dagger|\hat{b}_i\rangle\langle\hat{b}_i|\hat{U}_i) \\ &= \bigotimes_{i=1}^n \hat{\rho}(\mathcal{U}_i,\mathcal{E}_i,\hat{U}_i,\hat{b}_1). \end{aligned} \tag{37}$$

□

We conclude this section with a nontrivial generalization of Proposition S2 in [11], which shows that tensor product noise cannot affect the nice factorization properties of classical shadows with tensor product structure.

**Lemma 3.12.** *Let $0 \leq k \leq n$. Let $O \in \mathbb{H}_2^{\otimes n}$ be an $n$-qubit operator that acts nontrivially as $\widetilde{O} \in \mathbb{H}_2^{\otimes k}$ on $k$ qubits $i_1, \ldots, i_k$. Let $\mathcal{U} = \bigotimes_{i=1}^n \mathcal{U}_i$ be a product ensemble, and let $\mathcal{E} = \mathcal{E}_1 \otimes \ldots \otimes \mathcal{E}_n$ be a product channel. Assume that $\mathcal{M}_{\mathcal{U},\mathcal{E}}$ is invertible. If $k = 0$, then $\|O\|_{\text{shadow},\mathcal{U},\mathcal{E}} = 1$. Otherwise, if $k \geq 1$, then*

$$\|O\|_{\text{shadow},\mathcal{U},\mathcal{E}} = \left\|\widetilde{O}\right\|_{\text{shadow},\mathcal{U}_{i_1}\otimes\cdots\otimes\mathcal{U}_{i_k},\mathcal{E}_{i_1}\otimes\cdots\otimes\mathcal{E}_{i_k}}. \tag{38}$$

*Proof.* By Claim 3.10, given a product ensemble and product channel, the shadow channel and inverse shadow channel factorize. It follows that that $\mathcal{M}_{\mathcal{U},\mathcal{E}}^{-1,\dagger} = \bigotimes_{i=1}^n \mathcal{M}_{\mathcal{U}_i,\mathcal{E}_i}^{-1,\dagger}$. Additionally, the trace preserving property of $\mathcal{M}_{\mathcal{U}_i,\mathcal{E}_i}$ implies that $\mathcal{M}_{\mathcal{U}_i,\mathcal{E}_i}^{-1}$ is trace preserving. The complex conjugate of a trace-preserving quantum channel is unital (see [100], Theorem 2.26). Therefore, $\mathcal{M}_{\mathcal{U}_i,\mathcal{E}_i}^{-1,\dagger}$ is unital.

If $k = 0$, then $O = \mathbb{I}$. It follows from Footnote 6 that $\|\mathbb{I}\|_{\text{shadow},\mathcal{U},\mathcal{E}}^2 = 1$. Without loss of generality, take $O = \widetilde{O} \otimes \mathbb{I}^{\otimes(n-k)}$.

$$\begin{aligned}\|O\|_{\text{shadow},\mathcal{U},\mathcal{E}}^2 &= \left\|\widetilde{O} \otimes \mathbb{I}^{\otimes(n-k)}\right\|_{\text{shadow},\mathcal{U},\mathcal{E}}^2 \\ &= \max_{\sigma\in\mathbb{D}_{2^n}} \mathop{\mathbb{E}}_{U\sim\mathcal{U}} \sum_{b\in\{0,1\}^n} \langle b|\mathcal{E}(U\sigma U^\dagger)|b\rangle\langle b|U\mathcal{M}_{\mathcal{U},\mathcal{E}}^{-1,\dagger}(\widetilde{O}\otimes\mathbb{I}^{\otimes(n-k)})U^\dagger|b\rangle^2 \\ &= \max_{\sigma\in\mathbb{D}_{2^n}} \mathop{\mathbb{E}}_{U\sim\mathcal{U}} \sum_{b\in\{0,1\}^n} \langle b|\mathcal{E}(U\sigma U^\dagger)|b\rangle\langle b|U\left(\bigotimes_{i=1}^k \mathcal{M}_{\mathcal{U}_i,\mathcal{E}_i}^{-1,\dagger}(\widetilde{O})\otimes\mathbb{I}\right)U^\dagger|b\rangle^2. \end{aligned} \tag{39}$$

The last equality follows from the fact that $\mathcal{M}_{\mathcal{U},\mathcal{E}}^{-1,\dagger}$ factorizes and is unital. We write $U \in \mathcal{U} = \mathcal{U}_1 \otimes \ldots \otimes \mathcal{U}_n$ as $V \otimes W$, with $V = U_1 \otimes \ldots \otimes U_k$, $W = U_{k+1} \otimes \ldots \otimes U_n$. We also write $\mathcal{U}_1 \otimes \ldots \otimes \mathcal{U}_k$ and $\mathcal{U}_{k+1} \otimes \ldots \otimes \mathcal{U}_n$ as $\mathcal{U}_{1\ldots k}$ and $\mathcal{U}_{k+1\ldots n}$, respectively. The expression becomes

$$\begin{aligned}&= \max_{\sigma\in\mathbb{D}_{2^n}} \mathop{\mathbb{E}}_{V\sim\mathcal{U}_{1\ldots k}} \mathop{\mathbb{E}}_{W\sim\mathcal{U}_{k+1\ldots n}} \sum_{c\in\{0,1\}^k} \sum_{d\in\{0,1\}^{(n-k)}} \langle c|\langle d|\mathcal{E}(V\otimes W\sigma V^\dagger\otimes W^\dagger)|c\rangle|d\rangle \\ &\quad \cdot \langle c|\langle d|(V\otimes W)\left(\bigotimes_{i=1}^k \mathcal{M}_{\mathcal{U}_i,\mathcal{E}_i}^{-1,\dagger}(\widetilde{O})\otimes\mathbb{I}\right)(V^\dagger\otimes W^\dagger)|c\rangle|d\rangle^2\end{aligned}$$

$$
\begin{aligned}
&= \max_{\sigma \in \mathbb{D}_{2^n}} \mathop{\mathbb{E}}_{V \sim \mathcal{U}_{1\ldots k}} \mathop{\mathbb{E}}_{W \sim \mathcal{U}_{k+1\ldots n}} \sum_{c \in \{0,1\}^k} \sum_{d \in \{0,1\}^{(n-k)}} \langle c|\langle d|\mathcal{E}(V \otimes W\sigma V^\dagger \otimes W^\dagger)|c\rangle|d\rangle \\
&\qquad \cdot \langle c|V \bigotimes_{i=1}^{k} \mathcal{M}^{-1,\dagger}_{\mathcal{U}_i,\mathcal{E}_i}(\widetilde{O})V^\dagger|c\rangle^2 \\
&= \max_{\sigma \in \mathbb{D}_{2^n}} \mathop{\mathbb{E}}_{V \sim \mathcal{U}_{1\ldots k}} \sum_{c \in \{0,1\}^k} \langle c|V \bigotimes_{i=1}^{k} \mathcal{M}^{-1,\dagger}_{\mathcal{U}_i,\mathcal{E}_i}(\widetilde{O})V^\dagger|c\rangle^2 \\
&\qquad \cdot \langle c| \left( \mathop{\mathbb{E}}_{W \sim \mathcal{U}_{k+1\ldots n}} \sum_{d \in \{0,1\}^{(n-k)}} \langle d|\mathcal{E}(V \otimes W\sigma V^\dagger \otimes W^\dagger)|d\rangle \right) |c\rangle.
\end{aligned} \tag{40}
$$

To simplify the expression further, we focus on the summation over $d \in \{0,1\}^{n-k}$, which is exactly the partial trace over the last $n-k$ qubits. We denote the partial trace over the last $n-k$ qubits as $\mathrm{tr}_{k+1\ldots n}$. We write $\sigma = \sum_a E_a \otimes F_a$, where $E_a \in \mathcal{L}(\mathbb{C}^{2^n}), F_a \in \mathcal{L}(\mathbb{C}^{2^{n-k}})$, and $\mathcal{E} = \mathcal{E}_1 \otimes \ldots \otimes \mathcal{E}_n = \mathcal{E}_{1\ldots k} \otimes \mathcal{E}_{k+1\ldots n}$, where $\mathcal{E}_{1\ldots k} = \mathcal{E}_1 \otimes \ldots \otimes \mathcal{E}_k, \mathcal{E}_{k+1\ldots n} = \mathcal{E}_{k+1} \otimes \ldots \otimes \mathcal{E}_n$.

$$
\begin{aligned}
\sum_{d \in \{0,1\}^{(n-k)}} \langle d|\mathcal{E}(V \otimes W\sigma V^\dagger \otimes W^\dagger)|d\rangle &= \mathrm{tr}_{k+1\ldots n}\, \mathcal{E}(V \otimes W\sigma V^\dagger \otimes W^\dagger) \\
&= \sum_a \mathrm{tr}_{k+1\ldots n} \left( \mathcal{E}_{1\ldots k}(VE_aV^\dagger) \otimes \mathcal{E}_{k+1\ldots n}(WF_aW^\dagger) \right) \\
&= \sum_a \mathcal{E}_{1\ldots k}(VE_aV^\dagger)\, \mathrm{tr}\left( \mathcal{E}_{k+1\ldots n}(WF_aW^\dagger) \right) \\
&= \sum_a \mathcal{E}_{1\ldots k}(VE_aV^\dagger)\, \mathrm{tr}(WF_aW^\dagger) \\
&= \sum_a \mathcal{E}_{1\ldots k}(VE_aV^\dagger)\, \mathrm{tr}(F_a) \\
&= \mathcal{E}_{1\ldots k}(V \sum_a E_a\, \mathrm{tr}(F_a) V^\dagger) \\
&= \mathcal{E}_{1\ldots k}(V\, \mathrm{tr}_{k+1}(\sigma) V^\dagger).
\end{aligned} \tag{41}
$$

Plugging into the expression for the shadow seminorm, we get

$$
\begin{aligned}
&= \max_{\sigma \in \mathbb{D}_{2^n}} \mathop{\mathbb{E}}_{V \sim \mathcal{U}_{1\ldots k}} \sum_{c \in \{0,1\}^k} \langle c|V \bigotimes_{i=1}^{k} \mathcal{M}^{-1,\dagger}_{\mathcal{U}_i,\mathcal{E}_i}(\widetilde{O})V^\dagger|c\rangle^2 \langle c|\Big( \mathop{\mathbb{E}}_{W \sim \mathcal{U}_{k+1\ldots n}} \mathcal{E}_{1\ldots k}(V\, \mathrm{tr}_{k+1}(\sigma)V^\dagger) \Big)|c\rangle \\
&= \max_{\sigma \in \mathbb{D}_{2^n}} \mathop{\mathbb{E}}_{V \sim \mathcal{U}_{1\ldots k}} \sum_{c \in \{0,1\}^k} \langle c|\mathcal{E}_{1\ldots k}(V\, \mathrm{tr}_{k+1}(\sigma)V^\dagger)|c\rangle\langle c|V\mathcal{M}^{-1,\dagger}_{\mathcal{U}_{1\ldots k},\mathcal{E}_{1\ldots k}}(\widetilde{O})V^\dagger|c\rangle^2.
\end{aligned} \tag{42}
$$

Because the partial trace preserves the space of quantum states,

$$
\begin{aligned}
\|O\|^2_{\mathrm{shadow},\mathcal{U},\mathcal{E}} &= \max_{\tau \in \mathbb{D}_{2^k}} \mathop{\mathbb{E}}_{V \sim \mathcal{U}_{1\ldots k}} \sum_{c \in \{0,1\}^k} \langle c|\mathcal{E}_{1\ldots k}(V\tau V^\dagger)|c\rangle\langle c|V\mathcal{M}^{-1,\dagger}_{\mathcal{U}_{1\ldots k},\mathcal{E}_{1\ldots k}}(\widetilde{O})V^\dagger|c\rangle^2 \\
&= \left\|\widetilde{O}\right\|^2_{\mathrm{shadow},\mathcal{U}_{1\ldots k},\mathcal{E}_{1\ldots k}}.
\end{aligned} \tag{43}
$$

□

## 4 Global Clifford Ensemble with Noise

In this section we prove that if $\mathcal{U}$ is the Clifford group and $\mathcal{E}$ is an arbitrary quantum channel, the shadow channel can be expressed as a depolarizing channel (Definition 2.2). In this setting, we derive the expression for the classical shadow and the sample complexity of the classical shadows protocol.

### 4.1 Derivation of Shadow Channel

We begin with a technical lemma involving 2-design ensembles.

**Lemma 4.1.** *Let $\mathcal{U}$ be an $n$-qubit 2-design and $\mathcal{E} : \mathcal{L}(\mathbb{C}^{2^n}) \to \mathcal{L}(\mathbb{C}^{2^n})$ be a linear superoperator. Then,*

$$\mathcal{M}_{\mathcal{U},\mathcal{E}}(A) = f(\mathcal{E})A + \Big(\frac{1}{2^n}\operatorname{tr}(\mathcal{E}(\mathbb{I})) - f(\mathcal{E})\Big)\operatorname{tr}(A)\frac{\mathbb{I}}{2^n}, \tag{44}$$

*where*

$$f(\mathcal{E}) = \frac{1}{2^{2n}-1}\Big(\operatorname{tr}(\mathcal{E}\circ\operatorname{diag}) - \frac{1}{2^n}\operatorname{tr}(\mathcal{E}(\mathbb{I}))\Big). \tag{45}$$

*Also, if $\mathcal{E}$ is trace-preserving or unital, then,*

$$\mathcal{M}_{\mathcal{U},\mathcal{E}} = \mathcal{D}_{n,f(\mathcal{E})}, \qquad \textit{where} \quad f(\mathcal{E}) = \frac{\operatorname{tr}(\mathcal{E}\circ\operatorname{diag}) - 1}{2^{2n}-1}. \tag{46}$$

*Proof.* We first introduce the following notation: let $\mathcal{E} : \mathcal{L}(\mathbb{C}^d) \to L(\mathbb{C}^d)$ be a linear superoperator. Define the unary operator $()^\ddagger$ as follows:

$$\mathcal{E}^\ddagger(A) = (\mathcal{E}^*(A^\dagger))^\dagger. \tag{47}$$

Say that $\mathcal{E}$ has Kraus representation

$$\mathcal{E} : B \mapsto \sum_i J_i B K_i^\dagger. \tag{48}$$

Then, the shadow channel may be evaluated as

$$\begin{aligned}
\mathcal{M}_{\mathcal{U},\mathcal{E}}(A) &= \mathop{\mathbb{E}}_{U\sim\mathcal{U}} \sum_{b\in\{0,1\}^n} \langle b|\mathcal{E}(UAU^\dagger)|b\rangle U^\dagger|b\rangle\langle b|U \\
&= \mathop{\mathbb{E}}_{U\sim\mathcal{U}} \sum_{b\in\{0,1\}^n} \langle b|\left(\sum_i J_i UAU^\dagger K_i^\dagger\right)|b\rangle U^\dagger|b\rangle\langle b|U \\
&= \sum_{b\in\{0,1\}^n} \mathop{\mathbb{E}}_{U\sim\mathcal{U}} \operatorname{tr}\left(U^\dagger \sum_i K_i^\dagger|b\rangle\langle b|J_i UA\right) U^\dagger|b\rangle\langle b|U \\
&= \sum_{b\in\{0,1\}^n} \mathop{\mathbb{E}}_{U\sim\mathcal{U}} \operatorname{tr}_1\left(U^\dagger\mathcal{E}^\ddagger(|b\rangle\langle b|)UA \otimes U^\dagger|b\rangle\langle b|U\right) \\
&= \sum_{b\in\{0,1\}^n} \operatorname{tr}_1\left\{\mathop{\mathbb{E}}_{U\sim\mathcal{U}} (U^\dagger\otimes U^\dagger)(\mathcal{E}^\ddagger(|b\rangle\langle b|)\otimes|b\rangle\langle b|)(U\otimes U)(A\otimes I)\right\}
\end{aligned}$$

$$= \sum_{b\in\{0,1\}^n} \mathrm{tr}_1 \left\{ \mathop{\mathbb{E}}_{U\sim\mathcal{U}^\dagger} (U\otimes U)(\mathcal{E}^\ddagger(|b\rangle\langle b|)\otimes|b\rangle\langle b|)(U^\dagger\otimes U^\dagger)(A\otimes I) \right\}$$

$$= \sum_{b\in\{0,1\}^n} \mathrm{tr}_1 \left\{ \underbrace{T_2^{(2^n)}\left(\mathcal{E}^\ddagger(|b\rangle\langle b|)\otimes|b\rangle\langle b|\right)}_{1}(A\otimes I) \right\}, \tag{49}$$

since $\mathcal{U}$ is a 2-design. In the above equations, $\mathrm{tr}_1$ denotes the partial trace over the first subsystem.

Applying Eq. (21) to the 2-fold twirl 1 gives

$$1 = \int \mathrm{d}\eta(U)\ (U\otimes U)\left[\mathcal{E}^\ddagger(|b\rangle\langle b|)\otimes|b\rangle\langle b|\right](U^\dagger\otimes U^\dagger)$$

$$= \frac{1}{2^{2n}-1}\left[\mathrm{tr}\left(\mathcal{E}^\ddagger(|b\rangle\langle b|)\otimes|b\rangle\langle b|\right)\left(I-\frac{W}{2^n}\right)+\mathrm{tr}\left(W\mathcal{E}^\ddagger(|b\rangle\langle b|)\otimes|b\rangle\langle b|\right)\left(W-\frac{I}{2^n}\right)\right], \tag{50}$$

where the traces in the above equations simplify as

$$\mathrm{tr}\left(\mathcal{E}^\ddagger(|b\rangle\langle b|)\otimes|b\rangle\langle b|\right) = \mathrm{tr}(\mathcal{E}^\ddagger(|b\rangle\langle b|)), \tag{51}$$

$$\mathrm{tr}\left(W\mathcal{E}^\ddagger(|b\rangle\langle b|)\otimes|b\rangle\langle b|\right) = \langle b|\mathcal{E}^\ddagger(|b\rangle\langle b|)|b\rangle. \tag{52}$$

Therefore,

$$\mathcal{M}_{\mathcal{U},\mathcal{E}}(A) = \sum_{b\in\{0,1\}^n} \mathrm{tr}_1\left\{\frac{1}{2^{2n}-1}\left(\mathrm{tr}(\mathcal{E}^\ddagger(|b\rangle\langle b|))\left(I-\frac{W}{2^n}\right)+\langle b|\mathcal{E}^\ddagger(|b\rangle\langle b|)|b\rangle\left(W-\frac{I}{2^n}\right)\right)(A\otimes I)\right\}$$

$$= \frac{1}{2^{2n}-1}\Bigg[\underbrace{\mathrm{tr}(\mathcal{E}^\ddagger(I))}_{2}\underbrace{\mathrm{tr}_1\left\{\left(I-\frac{W}{2^n}\right)(A\otimes I)\right\}}_{3}$$

$$+ \underbrace{\sum_{b\in\{0,1\}^n}\langle b|\mathcal{E}^\ddagger(|b\rangle\langle b|)|b\rangle}_{4}\underbrace{\mathrm{tr}_1\left\{\left(W-\frac{I}{2^n}\right)(A\otimes I)\right\}}_{5}\Bigg]. \tag{53}$$

Then, by simple calculation,

$$2 = \mathrm{tr}(\mathcal{E}^\ddagger(I)) = \mathrm{tr}(\mathcal{E}(I)), \tag{54}$$

and

$$4 = \mathrm{tr}(\mathcal{E}^\ddagger\circ\mathrm{diag}) = \mathrm{tr}(\mathcal{E}\circ\mathrm{diag}). \tag{55}$$

To evaluate 3 and 5, we use the fact that

$$\mathrm{tr}_1(W(A\otimes I)) = A. \tag{56}$$

Hence,

$$\textcircled{3} = \mathrm{tr}_1(A \otimes I) - \frac{1}{2^n}\mathrm{tr}_1(W(A \otimes I))$$
$$= \mathrm{tr}(A)I - \frac{1}{2^n}A \tag{57}$$

and

$$\textcircled{5} = \mathrm{tr}_1(W(A \otimes I)) - \frac{1}{2^n}\mathrm{tr}_1(A \otimes I)$$
$$= A - \frac{1}{2^n}\mathrm{tr}(A)I. \tag{58}$$

Plugging these back into Eq. (53),

$$\mathcal{M}_{\mathcal{U},\mathcal{E}}(A) = \frac{1}{2^{2n}-1}\left[\mathrm{tr}(\mathcal{E}(I))(\mathrm{tr}(A)I - \frac{1}{2^n}A) + \mathrm{tr}(\mathcal{E} \circ \mathrm{diag})(A - \frac{1}{2^n}\mathrm{tr}(A)I)\right]$$
$$= f(\mathcal{E})A + \Big(\frac{1}{2^n}\mathrm{tr}(\mathcal{E}(\mathbb{I})) - f(\mathcal{E})\Big)\mathrm{tr}(A)\frac{\mathbb{I}}{2^n}, \tag{59}$$

where

$$f(\mathcal{E}) = \frac{1}{2^{2n}-1}\Big(\mathrm{tr}(\mathcal{E} \circ \mathrm{diag}) - \frac{1}{2^n}\mathrm{tr}(\mathcal{E}(\mathbb{I}))\Big), \tag{60}$$

which completes the first part of the proof. The second part of the proof follows from the fact that, if $\mathcal{E}$ is trace-preserving or unital, then $\mathrm{tr}(\mathcal{E}(\mathbb{I})) = \mathrm{tr}(\mathbb{I}) = 2^n$. Substituting this into Eq. (44) and Eq. (45) gives the result. □

Since the Clifford group forms a 2-design, Lemma 4.1 immediately implies that if $\mathcal{E}$ is an arbitrary quantum channel (and is hence trace-preserving), then the shadow channel is given by

$$\mathcal{M}_{\mathcal{C}_n,\mathcal{E}} = \mathcal{D}_{n,f(\mathcal{E})}, \qquad \text{where} \quad f(\mathcal{E}) = \frac{\mathrm{tr}(\mathcal{E} \circ \mathrm{diag}) - 1}{2^{2n}-1}. \tag{61}$$

Moreover, if the shadow channel is invertible, its inverse is given by $\mathcal{M}^{-1}_{\mathcal{C}_n,\mathcal{E}} = \mathcal{D}_{n,f(\mathcal{E})^{-1}}$.

An important fact can be deduced from Lemma 4.1, namely, that if the unitary ensemble is a 2-design, then the shadow channel is invertible if and only if the error channel obeys the simple condition $\mathrm{tr}(\mathcal{E} \circ \mathrm{diag}) \neq 1$:

**Claim 4.2.** *Let $\mathcal{U}$ be an $n$-qubit 2-design, and let $\mathcal{E}$ be a linear superoperator. $\mathcal{M}_{\mathcal{U},\mathcal{E}}$ is invertible if and only if $\mathrm{tr}(\mathcal{E} \circ \mathrm{diag}) \neq 1$. In this case, $\mathcal{M}^{-1}_{\mathcal{U},\mathcal{E}} = \mathcal{D}_{n,1/f(\mathcal{E})}$.*

*Proof.* By Lemma 4.1, the shadow channel with noise is a depolarizing channel. Therefore,

$$\mathcal{M}_{\mathcal{U},\mathcal{E}} = \mathcal{D}_{n,f(\mathcal{E})} \text{ is invertible} \iff f(\mathcal{E}) \neq 0 \iff \mathrm{tr}(\mathcal{E} \circ \mathrm{diag}) \neq 1, \tag{62}$$

and

$$\mathcal{M}^{-1}_{\mathcal{U},\mathcal{E}} = \mathcal{D}^{-1}_{n,f(\mathcal{E})} = \mathcal{D}_{n,1/f(\mathcal{E})}. \tag{63}$$

□

Next, we prove bounds on the depolarizing parameter $f(\mathcal{E})$:

**Claim 4.3.** *Let $\mathcal{E} : \mathcal{L}(\mathbb{C}^{2^n}) \to \mathcal{L}(\mathbb{C}^{2^n})$ be a quantum channel. Then,*

$$-\frac{1}{2^{2n}-1} \leq f(\mathcal{E}) \leq \frac{1}{2^n+1}. \tag{64}$$

*Proof.*

$$\begin{aligned}\mathcal{E} \text{ is a quantum channel} &\implies \mathrm{tr}(\mathcal{E} \circ \mathrm{diag}) \in [0, 2^n] \\ &\implies f(\mathcal{E}) = \frac{\mathrm{tr}(\mathcal{E} \circ \mathrm{diag}) - 1}{2^{2n}-1} \in \Big[-\frac{1}{2^{2n}-1}, \frac{1}{2^n+1}\Big].\end{aligned} \tag{65}$$

□

A few remarks are in order. First, note that the bounds in Claim 4.3 have appeared in work on randomized benchmarking (e.g., Lemma 1 in [108]).

Second, note that the depolarizing parameter $f(\mathcal{E})$ is upper bounded by $\frac{1}{2^n+1} = f(\mathbb{I})$, which is the depolarizing parameter of the noiseless shadow channel $\mathcal{M}_{\mathcal{U}}$. In other words, as expected, noise necessarily decreases the depolarizing parameter of the shadow channel (i.e., it is not possible to use noise to improve the performance of classical shadows).

Finally, note from Eq. (64) that $f(\mathcal{E})$ can take negative values. As discussed in Section 2.1, while it is typical to consider depolarizing channels with depolarizing parameter $f \in [0, 1]$, $D_{n,f}$ remains a quantum channel for some negative values of $f$. We note here that the lower bound in Eq. (64) matches exactly the lower bound in Eq. (13) when $d = 2^n$.

## 4.2 Classical Shadow

We now give an expression for the classical shadow when $\mathcal{M}_{\mathcal{U},\mathcal{E}} = \mathcal{D}_{n,f(\mathcal{E})}$. Recall that the classical shadow corresponding to a unitary ensemble $\mathcal{U}$, noise channel $\mathcal{E}$, unitary transformation $\hat{U} \in \mathcal{U}$, and bit string $\hat{b} \in \{0,1\}^n$ is $\hat{\rho}(\mathcal{U}, \mathcal{E}, \hat{U}, \hat{b}) = \mathcal{M}^{-1}_{\mathcal{U},\mathcal{E}}(\hat{U}^\dagger|\hat{b}\rangle\langle\hat{b}|\hat{U})$.

**Claim 4.4.** *Let $\mathcal{U}$ be an $n$-qubit 2-design, and let $\mathcal{E}$ be a quantum channel. Assume $\mathcal{M}_{\mathcal{U},\mathcal{E}}$ is invertible. Then, for some $\hat{U} \in \mathcal{U}$ and $\hat{b} \in \{0,1\}^n$, the classical shadow is*

$$\hat{\rho}(\mathcal{U}, \mathcal{E}, \hat{U}, \hat{b}) = \frac{1}{f(\mathcal{E})}\hat{U}|\hat{b}\rangle\langle\hat{b}| + \Big(1 - \frac{1}{f(\mathcal{E})}\Big)\frac{\mathbb{I}}{2^n}, \tag{66}$$

*where $f(\mathcal{E}) = \frac{\mathrm{tr}(\mathcal{E}\circ\mathrm{diag})-1}{2^{2n}-1}$.*

*Proof.*

$$\begin{aligned}\hat{\rho}(\mathcal{U}, \mathcal{E}, \hat{U}, \hat{b}) &= \mathcal{M}^{-1}_{\mathcal{U},\mathcal{E}}(\hat{U}^\dagger|\hat{b}\rangle\langle\hat{b}|\hat{U}) \\ &= \mathcal{D}^{-1}_{n,f(\mathcal{E})}(\hat{U}^\dagger|\hat{b}\rangle\langle\hat{b}|\hat{U}) \\ &= \mathcal{D}_{n,1/f(\mathcal{E})}(\hat{U}^\dagger|\hat{b}\rangle\langle\hat{b}|\hat{U}) \\ &= \frac{1}{f(\mathcal{E})}\hat{U}|\hat{b}\rangle\langle\hat{b}|\hat{U} + \Big(1 - \frac{1}{f(\mathcal{E})}\Big)\frac{\mathbb{I}}{2^n}.\end{aligned} \tag{67}$$

□

### 4.3 Derivation of Shadow Seminorm

We derive an expression for the shadow seminorm of a traceless observable. Recall that the sample complexity to estimate $\mathrm{tr}(O\rho)$ for some observable $O$ is upper bounded by the shadow seminorm of the traceless part of $O$ (Lemma 3.6), which we write as $O_o$.

**Proposition 4.5.** *Let $\mathcal{U}$ be an $n$-qubit 3-design, and let $\mathcal{E}$ be a linear superoperator, and let $O_o$ be a traceless observable. Then,*

$$\|O_o\|^2_{\mathrm{shadow},\mathcal{U},\mathcal{E}} = \frac{d(d^2-1)}{(d+2)(d\beta-\alpha)}\left(\frac{(1+d)\alpha-2\beta}{d\beta-\alpha}\,\mathrm{tr}(O_o^2) + 2\left\|O_o^2\right\|_{\mathrm{sp}}\right) \tag{68}$$

*where*

$$d = 2^n, \quad \alpha = \mathrm{tr}(\mathcal{E}(\mathbb{I})), \quad \beta = \mathrm{tr}(\mathcal{E}\circ\mathrm{diag}). \tag{69}$$

*Also, if $\mathcal{E}$ be a trace-preserving or unital linear superoperator, then the expression simplifies to*

$$\|O_o\|^2_{\mathrm{shadow},\mathcal{U},\mathcal{E}} = \frac{d^2-1}{(d+2)(\beta-1)}\left(\frac{d+d^2-2\beta}{d(\beta-1)}\,\mathrm{tr}(O_o^2) + 2\left\|O_o^2\right\|_{\mathrm{sp}}\right) \tag{70}$$

*where*

$$d = 2^n, \qquad \beta = \mathrm{tr}(\mathcal{E}\circ\mathrm{diag}). \tag{71}$$

*Proof.*

$$\begin{aligned}
\|O_o\|^2_{\mathrm{shadow},\mathcal{U},\mathcal{E}} &= \max_{\sigma\in\mathbb{D}_{2^n}} \mathop{\mathbb{E}}_{U\sim\mathcal{U}} \sum_{b\in\{0,1\}^n} \langle b|\mathcal{E}(U\sigma U^\dagger)|b\rangle\langle b|U\mathcal{M}^{-1,\dagger}_{\mathcal{U},\mathcal{E}}(O_o)U^\dagger|b\rangle^2 \\
&= \max_{\sigma\in\mathbb{D}_{2^n}} \mathop{\mathbb{E}}_{U\sim\mathcal{U}} \sum_{b\in\{0,1\}^n} \langle b|\mathcal{E}(U\sigma U^\dagger)|b\rangle\langle b|U\mathcal{D}_{n,1/f(\mathcal{E})}(O_o)U^\dagger|b\rangle^2 \\
&= \frac{1}{f(\mathcal{E})^2}\max_{\sigma\in\mathbb{D}_{2^n}} \mathop{\mathbb{E}}_{U\sim\mathcal{U}} \sum_{b\in\{0,1\}^n} \langle b|\mathcal{E}(U\sigma U^\dagger)|b\rangle\langle b|UO_oU^\dagger|b\rangle^2 \\
&= \frac{1}{f(\mathcal{E})^2}\left(\frac{(1+d)\alpha-2\beta}{(d-1)d(d+1)(d+2)}\,\mathrm{tr}(O_o^2) + \frac{2(d\beta-\alpha)}{(d-1)d(d+1)(d+2)}\left\|O_o^2\right\|_{\mathrm{sp}}\right).
\end{aligned} \tag{72}$$

The final equality follows from Lemma 2.6. To simplify further, we get an expression for $1/f(\mathcal{E})^2$.

$$\begin{aligned}
f(\mathcal{E}) &= \frac{1}{2^{2n}-1}\left(\mathrm{tr}(\mathcal{E}\circ\mathrm{diag}) - \frac{1}{2^n}\mathrm{tr}(\mathcal{E}(\mathbb{I}))\right) \\
&= \frac{d\beta-\alpha}{(d-1)d(d+1)} \\
\implies \frac{1}{f(\mathcal{E})^2} &= \frac{(d-1)^2d^2(d+1)^2}{(d\beta-\alpha)^2}.
\end{aligned} \tag{73}$$

Plugging into the expression above, we get

$$
\begin{aligned}
&\|O_o\|^2_{\mathrm{shadow},\mathcal{U},\mathcal{E}} \\
&= \frac{(d-1)^2 d^2 (d+1)^2}{(d\beta-\alpha)^2}\left(\frac{(1+d)\alpha - 2\beta}{(d-1)d(d+1)(d+2)}\operatorname{tr}(O_o^2) + \frac{2(d\beta-\alpha)}{(d-1)d(d+1)(d+2)}\left\|O_o^2\right\|_{\mathrm{sp}}\right) \\
&= \frac{d(d^2-1)}{d+2}\cdot\frac{(1+d)\alpha-2\beta}{(d\beta-\alpha)^2}\operatorname{tr}(O_o^2) + \frac{d(d^2-1)}{d+2}\cdot\frac{2}{d\beta-\alpha}\left\|O_o^2\right\|_{\mathrm{sp}} \\
&= \frac{d(d^2-1)}{(d+2)(d\beta-\alpha)}\left(\frac{(1+d)\alpha-2\beta}{d\beta-\alpha}\operatorname{tr}(O_o^2) + 2\left\|O_o^2\right\|_{\mathrm{sp}}\right).
\end{aligned}
\tag{74}
$$

Finally, if $\mathcal{E}$ is trace-preserving or unital, then $\alpha = \operatorname{tr}(\mathcal{E}(\mathbb{I})) = d$. The second part of the proposition follows from substituting $\alpha = d$. □

Building from Proposition 4.5, one can get looser bounds that are more convenient to work with.

**Corollary 4.6.** *Let $\mathcal{U}$ be an $n$-qubit 3-design, let $\mathcal{E}$ be a trace-preserving or unital linear superoperator, and let $O_o$ be a traceless observable. Then,*

$$
\frac{(2^n-1)^2}{(\beta-1)^2}\operatorname{tr}(O_o^2) \le \|O_o\|_{\mathrm{shadow},\mathcal{U},\mathcal{E}} \le \frac{3(2^n-1)^2}{(\beta-1)^2}\operatorname{tr}(O_o^2) \le \frac{3(2^n-1)^2}{(\beta-1)^2}\operatorname{tr}(O^2). \tag{75}
$$

*where $\beta = \operatorname{tr}(\mathcal{E}\circ\operatorname{diag})$.*

The proof is straightforward. We include it in Appendix A.3 for completeness.

Combining Theorem 3.7 and Corollary 4.6 yields sample complexity bounds on the classical shadows protocol when the unitary ensemble is the Clifford group (or, any unitary 3-design).

**Corollary 4.7.** *Let $\{O_i\}_{i=1}^M$ be a collection of $M$ observables. Let $\mathcal{U}$ be a unitary 3-design (e.g., the Clifford group). Let $\mathcal{E}:\mathcal{L}(\mathbb{C}^{2^n})\to\mathcal{L}(\mathbb{C}^{2^n})$ be a quantum channel such that $\operatorname{tr}(\mathcal{E}\circ\operatorname{diag})\neq 1$. The sample complexity $N_{\mathrm{tot}}$ to estimate the linear target functions $\{\operatorname{tr}(O_i\rho)\}_{i=1}^M$ of an $n$-qubit state $\rho$ within error $\varepsilon$ and failure probability $\delta$ when the unitary ensemble $\mathcal{U}$ is subject to the error channel $\mathcal{E}$ is*

$$
N_{\mathrm{tot}} \le \frac{204(2^n-1)^2\log(2M/\delta)}{(\beta-1)^2\varepsilon^2}\max_{1\le i\le M}\operatorname{tr}(O_i^2),
$$

*where $\beta = \operatorname{tr}(\mathcal{E}\circ\operatorname{diag})$.*

## 4.4 Examples

Eq. (61) establishes that if $\mathcal{U}$ is the Clifford group and $\mathcal{E}$ is an arbitrary quantum channel, then the resulting shadow channel $\mathcal{M}_{\mathcal{C}_n,\mathcal{E}}$ is always a depolarizing channel (Definition 2.2) with a depolarizing parameter that depends on the quantum channel $\mathcal{E}$. Corollary 4.7 establishes the sample complexity in this scenario. We now apply our results to derive expressions for the shadow channel, inverse shadow channel, classical shadow, shadow seminorm, and sample complexity for the classical shadows protocol with the Clifford group and specific quantum channels.

### 4.4.1 Noiseless Case

We begin with a basic example, the case where the quantum channel is the identity channel, to show that the results from [11] can be recovered. For reference, see Eqs. (S37) through (S43) in [11]).

**Claim 4.8.** *In the noiseless case,*

1. $f(\mathbb{I}) = \frac{1}{2^n+1}$.
2. $\mathcal{M}_{\mathcal{C}_n,\mathbb{I}} = \mathcal{D}_{n,1/2^n+1}$.
3. $\mathcal{M}^{-1}_{\mathcal{C}_n,\mathbb{I}} = \mathcal{D}_{n,2^n+1}$.
4. *The classical shadow can be written as* $\hat{\rho}(\mathcal{C}_n, \mathbb{I}, \hat{U}, \hat{b}) = (2^n+1)\hat{U}^\dagger|\hat{b}\rangle\langle\hat{b}|\hat{U} - \mathbb{I}$.
5. $\|O_o\|^2_{\text{shadow},\mathcal{C}_n,\mathbb{I}} = \frac{2^n+1}{2^n+2}\left(\operatorname{tr}(O_o^2) + 2\left\|O_o^2\right\|_{\text{sp}}\right)$.

*Proof.* (1) follows from a simple calculation:

$$f(\mathbb{I}) = \frac{\operatorname{tr}(\mathbb{I}\circ\text{diag}) - 1}{2^{2n}-1} = \frac{2^n-1}{2^{2n}-1} = \frac{1}{2^n+1}. \tag{76}$$

(2) and (3) follow from Eq. (61). To prove (4), apply Claim 4.4:

$$\begin{aligned}\hat{\rho}(\mathcal{C}_n, \mathbb{I}, \hat{U}, \hat{b}) &= (2^n+1)\hat{U}^\dagger|\hat{b}\rangle\langle\hat{b}|\hat{U} + (1 - 2^n + 1)\frac{\mathbb{I}}{2^n} \\ &= (2^n+1)\hat{U}^\dagger|\hat{b}\rangle\langle\hat{b}|\hat{U} - \mathbb{I}.\end{aligned} \tag{77}$$

To prove (5), apply Proposition 4.5 with $\beta = \operatorname{tr}(\mathbb{I}\circ\text{diag}) = 2^n = d$. Then,

$$\begin{aligned}\|O_o\|^2_{\text{shadow},\mathcal{C}_n,\mathbb{I}} &= \frac{d^2-1}{(d+2)(d-1)}\left(\frac{d+d^2-2d}{d(d-1)}\operatorname{tr}(O_o^2) + 2\left\|O_o^2\right\|_{\text{sp}}\right) \\ &= \frac{2^n+1}{2^n+2}\left(\operatorname{tr}(O_o^2) + 2\left\|O_o^2\right\|_{\text{sp}}\right).\end{aligned} \tag{78}$$

□

*Remark* 4.9. One can verify that the dephasing channel is an inconsequential noise channel (see Claim C.1 of Appendix C). As such, these results also hold when $\mathcal{E}$ is the dephasing channel.

### 4.4.2 Depolarizing Channel

We derive expressions for the shadow channel, inverse shadow channel, the classical shadow, and the shadow seminorm when the Clifford group $\mathcal{C}_n$ is subject to depolarizing noise with depolarizing parameter $f$ (Definition 2.2).

**Claim 4.10.** *If the unitary ensemble used in the classical shadows protocol is the Clifford group and is subject to depolarizing noise with depolarizing parameter* $f \in [0,1]$*, then*

1. $f(\mathcal{D}_{n,f}) = \frac{f}{2^n+1}$.

*2.* $\mathcal{M}_{\mathcal{C}_n,\mathcal{D}_{n,f}} = \mathcal{D}_{n,f/2^n+1}$.

*3.* $\mathcal{M}^{-1}_{\mathcal{C}_n,\mathcal{D}_{n,f}} = \mathcal{D}_{n,2^n+1/f}$.

*4. The classical shadow can be written as* $\hat{\rho}(\mathcal{C}_n,\mathcal{D}_{n,f},\hat{U},\hat{b}) = \frac{2^n+1}{f}\hat{U}^\dagger|\hat{b}\rangle\langle\hat{b}|\hat{U} - (1-\frac{2^n+1}{f})\frac{\mathbb{I}}{2^n}$.

*5. Let* $O \in \mathbb{H}_{2^n}$*. Then,* $\left\|O - \frac{1}{2^n}\operatorname{tr}(O)\mathbb{I}\right\|^2_{\text{shadow},\mathcal{C}_n,\mathcal{D}_{n,f}} \leq \frac{3}{f^2}\operatorname{tr}(O^2)$.

*Proof.* First we prove (1):

$$\begin{aligned}
\operatorname{tr}(\mathcal{D}_{n,f}\circ \operatorname{diag}) &= \sum_{b\in\{0,1\}^n} \langle b|\mathcal{D}_{n,f}(|b\rangle\langle b|)|b\rangle \\
&= \sum_{b\in\{0,1\}^n} \langle b|(f|b\rangle\langle b| + (1-f)\frac{\mathbb{I}}{2^n})|b\rangle \\
&= \sum_{b\in\{0,1\}^n} f + (1-f)\frac{1}{2^n} \\
&= 2^n f + 1 - f. \qquad (79)
\end{aligned}$$

Then,

$$\begin{aligned}
f(\mathcal{E}) &= \frac{\operatorname{tr}(\mathcal{D}_{n,f}\circ\operatorname{diag}) - 1}{2^{2n}-1} \\
&= \frac{2^n f + 1 - f - 1}{2^{2n}-1} \\
&= \frac{f(2^n-1)}{2^{2n}-1} \\
&= \frac{f}{2^n+1}. \qquad (80)
\end{aligned}$$

(2) and (3) follow from Eq. (61), and (5) follows from Corollary 4.6. To prove (4), apply Claim 4.4:

$$\hat{\rho}(\mathcal{C}_n,\mathbb{I},\hat{U},\hat{b}) = \frac{2^n+1}{f}\hat{U}^\dagger|\hat{b}\rangle\langle\hat{b}|\hat{U} - \Big(1-\frac{2^n+1}{f}\Big)\frac{\mathbb{I}}{2^n}. \qquad (81)$$

□

With a bound on the shadow seminorm in this setting, the sample complexity of the protocol follows.

**Corollary 4.11.** *The sample complexity* $N_{\text{tot}}$ *to estimate a collection of* $M$ *linear target functions* $\operatorname{tr}(O_i\rho)$ *within error* $\varepsilon$ *and failure probability* $\delta$ *when the unitary ensemble* $\mathcal{U}$ *is subject to depolarizing noise* $\mathcal{D}_{n,f} : A \mapsto fA + (1-f)\frac{\mathbb{I}}{2^n}$ *is*

$$N_{\text{tot}} \leq \frac{204\log(2M/\delta)}{f^2\varepsilon^2}\max_{1\leq i\leq M}\operatorname{tr}(O_i^2).$$

### 4.4.3 Amplitude Damping Channel

We derive expressions for the shadow channel, inverse shadow channel, the classical shadow, and the shadow seminorm when the Clifford group $\mathcal{C}_n$ is subject to the amplitude damping channel (Definition 2.3).

**Claim 4.12.** *If the unitary ensemble used in the classical shadows protocol is the Clifford group and is subject to amplitude damping noise with parameter $p \in [0,1]$, then*

1. $f(\mathrm{AD}_{n,p}) = \frac{(1+p)^n-1}{2^{2n}+1}$.
2. $\mathcal{M}_{\mathcal{C}_n,\mathrm{AD}_{n,p}} = \mathcal{D}_{n,((1+p)^n-1)/(2^{2n}-1)}$.
3. $\mathcal{M}^{-1}_{\mathcal{C}_n,\mathrm{AD}_{n,p}} = \mathcal{D}_{n,(2^{2n}-1)/((1+p)^n-1)}$.
4. *The classical shadow can be written as* $\hat{\rho}(\mathcal{C}_n, \mathrm{AD}_{n,p}, \hat{U}, \hat{b}) = \frac{2^n+1}{f}\hat{U}^\dagger|\hat{b}\rangle\langle\hat{b}|\hat{U} - (1-\frac{2^n+1}{f})\frac{\mathbb{I}}{2^n}$.
5. *Let $O \in \mathbb{H}_{2^n}$. Then,* $\left\|O - \frac{1}{2^n}\operatorname{tr}(O)\mathbb{I}\right\|^2_{\mathrm{shadow},\mathcal{C}_n,\mathrm{AD}_{n,p}} \leq \frac{3(2^n-1)^2}{((1+p)^n-1)^2}\operatorname{tr}(O^2)$.

*Proof.* To prove (1), we use the fact that $\langle 0|\mathrm{AD}_{1,p}(|0\rangle\langle 0|)|0\rangle = 1$ and $\langle 1|\mathrm{AD}_{1,p}(|1\rangle\langle 1|)|1\rangle = p$. We denote the Hamming weight of a bit string $b$ as $\ell_1(b)$.

$$\begin{aligned}
\operatorname{tr}(\mathrm{AD}_{n,p} \circ \mathrm{diag}) &= \sum_{b\in\{0,1\}^n} \langle b|\mathrm{AD}_{n,p}(|b\rangle\langle b|)|b\rangle \\
&= \sum_{b\in\{0,1\}^n} \prod_{i=1}^n \langle b_i|\mathrm{AD}_{1,p}(|b_i\rangle\langle b_i|)|b_i\rangle \\
&= \sum_{b\in\{0,1\}^n} p^{\ell_1(b)} \\
&= \sum_{i=0}^n \binom{n}{i} p^i \\
&= (1+p)^n.
\end{aligned} \tag{82}$$

Then,

$$\begin{aligned}
f(\mathrm{AD}_{n,p}) &= \frac{\operatorname{tr}(\mathrm{AD}_{n,p}\circ \mathrm{diag}) - 1}{2^{2n}-1} \\
&= \frac{(1+p)^n - 1}{2^{2n}-1}.
\end{aligned} \tag{83}$$

(2) and (3) follow from Eq. (61), and (5) follows from Corollary 4.6. To prove (4), apply Claim 4.4,

$$\begin{aligned}
\hat{\rho}(\mathcal{C}_n, \mathrm{AD}_{n,p}, \hat{U}, \hat{b}) &= \mathcal{D}_{n,(2^{2n}-1)/((1+p)^n-1)}(\hat{U}^\dagger|\hat{b}\rangle\langle\hat{b}|\hat{U}) \\
&= \frac{2^{2n}-1}{(1+p)^n-1}\hat{U}^\dagger|\hat{b}\rangle\langle\hat{b}|\hat{U} - \Big(1 - \frac{2^{2n}-1}{(1+p)^n-1}\Big)\frac{\mathbb{I}}{2^n}.
\end{aligned} \tag{84}$$

$\square$

A bound on the sample complexity follows from the bound on the shadow seminorm.

**Corollary 4.13.** *The sample complexity $N_{\rm tot}$ to estimate a collection of $M$ linear target functions $\operatorname{tr}(O_i\rho)$ within error $\varepsilon$ and failure probability $\delta$ when the unitary ensemble $\mathcal{U}$ is subject to the amplitude damping channel $\mathrm{AD}_{n,p}$ is*

$$N_{\rm tot} \leq \frac{204(2^n-1)^2\log(2M/\delta)}{((1+p)^n-1)^2\varepsilon^2}\max_{1\leq i\leq M}\operatorname{tr}(O_i^2).$$

# 5 Product Clifford Ensemble with Product Noise

In this section we analyze the setting in which the quantum channel is a product channel (Definition 3.8) and the unitary ensemble is a product ensemble (Definition 3.9). Our results hold for any product ensemble in which each ensemble is a 3-design, which the product Clifford ensemble is an example. We write the product Clifford ensemble as $\mathcal{C}_1^{\otimes n} = \{U_1\otimes\ldots\otimes U_n : U_1,\ldots,U_n\in\mathcal{C}_1\}$ and the quantum channel as $\mathcal{E}^{\otimes n}$, where $\mathcal{E}$ acts on a single qubit.

## 5.1 Derivation of Shadow Channel

We derive expressions for the shadow channel and its inverse by building on the work in Section 3.3 and Section 4.

**Claim 5.1.** *Let $\mathcal{U}=\mathcal{U}_1\otimes\ldots\otimes\mathcal{U}_n$ be a product ensemble such that $\mathcal{U}_i^\dagger$ is a 2-design for all $i\in[n]$. Let $\mathcal{E}:\mathcal{L}(\mathbb{C}^{2^n})\to\mathcal{L}(\mathbb{C}^{2^n})$ be a single-qubit quantum channel. Then,*

$$\mathcal{M}_{\mathcal{U},\mathcal{E}^{\otimes n}} = \mathcal{D}^{\otimes n}_{1,\frac{1}{3}(\operatorname{tr}(\mathcal{E}\circ\mathrm{diag})-1)}. \tag{85}$$

*Also, if* $\operatorname{tr}(\mathcal{E}\circ\mathrm{diag})\neq 1$, *then*

$$\mathcal{M}^{-1}_{\mathcal{U},\mathcal{E}^{\otimes n}} = \mathcal{D}^{\otimes n}_{1,3/(\operatorname{tr}(\mathcal{E}\circ\mathrm{diag})-1)}. \tag{86}$$

*Proof.* The first part follows from Claim 3.10 and Lemma 4.1. The second part follows from the fact that if $\operatorname{tr}(\mathcal{E}\circ\mathrm{diag})\neq 1$, then the shadow channel $\mathcal{M}_{\mathcal{U},\mathcal{E}^{\otimes n}}$ is invertible (see Claim 4.2). Therefore, $\mathcal{M}^{-1}_{\mathcal{U},\mathcal{E}^{\otimes n}} = (\mathcal{D}^{-1}_{1,(1/3)\cdot(\operatorname{tr}(\mathcal{E}\circ\mathrm{diag})-1)})^{\otimes n} = \mathcal{D}_{1,3/(\operatorname{tr}(\mathcal{E}\circ\mathrm{diag})-1)}$. ☐

## 5.2 Classical Shadow

We derive an expression for the classical shadow for the product Clifford ensemble and a product channel.

**Claim 5.2.** *Let $\mathcal{U}=\mathcal{U}_1\otimes\ldots\otimes\mathcal{U}_n$ be a product ensemble such that $\mathcal{U}_i^\dagger$ is a 2-design for all $i\in[n]$. Let $\mathcal{E}:\mathcal{L}(\mathbb{C}^{2^n})\to\mathcal{L}(\mathbb{C}^{2^n})$ be a single-qubit quantum channel such that $\operatorname{tr}(\mathcal{E}\circ\mathrm{diag})\neq 1$. Let $\hat{U}=\hat{U}_1\otimes\ldots\otimes\hat{U}_n\in\mathcal{U}$ and $\hat{b}=\hat{b}_1\ldots\hat{b}_n\in\{0,1\}^n$. Then,*

$$\hat{\rho}(\mathcal{U},\mathcal{E}^{\otimes n},\hat{U},\hat{b}) = \bigotimes_{i=1}^{n}\Big(\frac{1}{f(\mathcal{E})}\hat{U}|\hat{b}\rangle\langle\hat{b}|\hat{U} + (1-\frac{1}{f(\mathcal{E})})\frac{\mathbb{I}}{2}\Big). \tag{87}$$

*Proof.* Follows from Claim 4.4 by setting $n=1$. ☐

### 5.3 Derivation of Shadow Seminorm

We derive an expression for the shadow seminorm when the observable is a $k$-local Pauli observable, which non-trivially generalizes the Lemma S3 in [11]. Denote the set of Pauli operators by $\mathcal{P}_n = \{P_1 \otimes \ldots \otimes P_n : P_i \in \{\mathbb{I}, X, Y, Z\}\, \forall i \in \{1, \ldots, n\}\}$.

**Definition 5.3** (weight of Pauli operator)**.** The *weight* of the Pauli operator $P = P_1 \otimes \ldots \otimes P_n \in \mathcal{P}_n$ is $\mathrm{wt}(P) = |\{i : P_i \neq \mathbb{I}\}|$.

**Proposition 5.4.** *Let $\mathcal{U} = \mathcal{U}_1 \otimes \ldots \otimes \mathcal{U}_n$ be a product ensemble such that $\mathcal{U}_i$ is a 3-design for all $i \in [n]$. Let $\mathcal{E} : \mathcal{L}(\mathbb{C}^{2^n}) \to \mathcal{L}(\mathbb{C}^{2^n})$ be a single-qubit quantum channel such that $\mathrm{tr}(\mathcal{E} \circ \mathrm{diag}) \neq 1$. Let $P \in \mathcal{P}_n$. Then,*

$$\|P\|_{\mathrm{shadow},\mathcal{U},\mathcal{E}^{\otimes n}} = \Big(\frac{1}{\sqrt{3} f(\mathcal{E})}\Big)^{\mathrm{wt}(P)}. \tag{88}$$

*where $f(\mathcal{E}) = \frac{1}{3}(\mathrm{tr}(\mathcal{E} \circ \mathrm{diag}) - 1)$.*

*Proof.* Without loss of generality, write $P = P_1 \otimes \ldots \otimes P_k \otimes \mathbb{I}^{\otimes(n-k)}$. Then,

$$\begin{aligned}
\|P\|^2_{\mathrm{shadow},\mathcal{U},\mathcal{E}^{\otimes n}} &= \Big\|P_1 \otimes \ldots \otimes P_k \otimes \mathbb{I}^{\otimes(n-k)}\Big\|^2_{\mathrm{shadow},\mathcal{U},\mathcal{E}^{\otimes n}} \\
&= \|P_1 \otimes \ldots \otimes P_k\|^2_{\mathrm{shadow},\mathcal{U},\mathcal{E}^{\otimes n}} \\
&= \max_{\sigma \in \mathbb{D}_{2^n}} \mathop{\mathbb{E}}_{U \sim \mathcal{C}_1^{\otimes k}} \sum_{b \in \{0,1\}^k} \langle b|\mathcal{E}^{\otimes k}(U\sigma U^\dagger)|b\rangle\langle b|U \bigotimes_{i=1}^{k} \mathcal{D}_{1,f(\mathcal{E})^{-1}}(P_i) U^\dagger|b\rangle^2 \\
&= \max_{\sigma \in \mathbb{D}_{2^n}} \mathop{\mathbb{E}}_{U \sim \mathcal{C}_1^{\otimes k}} \sum_{b \in \{0,1\}^k} \langle b|\mathcal{E}^{\otimes k}(U\sigma U^\dagger)|b\rangle\langle b|U \bigotimes_{i=1}^{k} \Big(\frac{1}{f(\mathcal{E})} P_i\Big) U^\dagger|b\rangle^2 \\
&= \frac{1}{f(\mathcal{E})^{2k}} \max_{\sigma \in \mathbb{D}_{2^n}} \mathop{\mathbb{E}}_{U \sim \mathcal{C}_1^{\otimes k}} \sum_{b \in \{0,1\}^k} \langle b|\mathcal{E}^{\otimes k}(U\sigma U^\dagger)|b\rangle\langle b|U \bigotimes_{i=1}^{k} P_i U^\dagger|b\rangle^2.
\end{aligned} \tag{89}$$

The second equality follows from Lemma 3.12. To simplify the expression further, we write $\sigma = \sum_{\alpha,\beta \in \{0,1\}^k} \sigma_{\alpha\beta}|\alpha\rangle\langle\beta| = \sum_{\alpha,\beta \in \{0,1\}^k} \sigma_{\alpha\beta} E_{\alpha_1\beta_1} \otimes \ldots \otimes E_{\alpha_k\beta_k}$, where $E_{\alpha_i\beta_i} = |\alpha_i\rangle\langle\beta_i|$. We simplify the expectation value first.

$$\begin{aligned}
&\mathop{\mathbb{E}}_{U \sim \mathcal{C}_1^{\otimes k}} \sum_{b \in \{0,1\}^k} \langle b|\mathcal{E}^{\otimes k}(U\sigma U^\dagger)|b\rangle\langle b|U \bigotimes_{i=1}^{k} P_i U^\dagger|b\rangle^2 \\
&\quad= \mathop{\mathbb{E}}_{U \sim \mathcal{C}_1^{\otimes k}} \sum_{b \in \{0,1\}^k} \langle b|\mathcal{E}^{\otimes k}(U(\sum_{\alpha,\beta \in \{0,1\}^k} \sigma_{\alpha\beta} E_{\alpha_1\beta_1} \otimes \ldots \otimes E_{\alpha_k\beta_k})U^\dagger)|b\rangle\langle b|U \bigotimes_{i=1}^{k} P_i U^\dagger|b\rangle^2 \\
&\quad= \sum_{\alpha,\beta \in \{0,1\}^k} \sigma_{\alpha\beta} \prod_{i=1}^{k} \mathop{\mathbb{E}}_{U_j \sim \mathcal{C}_1} \sum_{b_j \in \{0,1\}} \langle b_j|\mathcal{E}(U_j E_{\alpha_j\beta_j} U^\dagger)|b_j\rangle\langle b_j|U_j P_j U_j^\dagger|b_j\rangle^2 \\
&\quad= \sum_{\alpha,\beta \in \{0,1\}^k} \sigma_{\alpha\beta} \prod_{i=1}^{k} \frac{1}{4!}\Big(2(6 - 2\,\mathrm{tr}(\mathcal{E} \circ \mathrm{diag})) + 2(2\,\mathrm{tr}(\mathcal{E} \circ \mathrm{diag}) - 2)\Big)\, \mathrm{tr}(E_{\alpha_j\beta_j})
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{\alpha,\beta\in\{0,1\}^k} \sigma_{\alpha\beta} \prod_{i=1}^{k} \frac{1}{3}\delta_{\alpha_j\beta_j} \\
&= \frac{1}{3^k}. \qquad (90)
\end{aligned}
$$

The third equality follows from Lemma 2.6 with $B = C = P_j$, $d = 2$ and $A = E_{\alpha_j\beta_j}$. Plugging into the original expression, we get

$$
\begin{aligned}
\|P\|^2_{\text{shadow},\mathcal{U},\mathcal{E}^{\otimes n}} &= \frac{1}{f(\mathcal{E})^{2k}} \max_{\sigma\in\mathbb{D}_{2^n}} \mathop{\mathbb{E}}_{U\sim\mathcal{C}_1^{\otimes k}} \sum_{b\in\{0,1\}^k} \langle b|\mathcal{E}^{\otimes k}(U\sigma U^\dagger)|b\rangle\langle b|U \bigotimes_{i=1}^{k} P_i U^\dagger|b\rangle^2 \\
&= \Big(\frac{1}{\sqrt{3}f(\mathcal{E})}\Big)^{2k} \qquad (91) \\
\implies \|P\|_{\text{shadow},\mathcal{U},\mathcal{E}^{\otimes n}} &= \Big(\frac{1}{\sqrt{3}f(\mathcal{E})}\Big)^{\text{wt}(P)}. \qquad (92)
\end{aligned}
$$

□

Since we have a bound on the shadow seminorm, we can bound the sample complexity of classical shadows protocol when the unitary ensemble is the product Clifford group and when all the observables are $k$-local Pauli operators.

**Corollary 5.5.** *Let $\{P_i\}_{i=1}^M$ be a collection of $M$ Pauli operators. Let $\mathcal{U} = \mathcal{U}_1 \otimes \ldots \otimes \mathcal{U}_n$ be a product ensemble such that $\mathcal{U}_i^\dagger$ is a 3-design for all $i \in [n]$. Let $\mathcal{E} : \mathcal{L}(\mathbb{C}^2) \to \mathcal{L}(\mathbb{C}^2)$ be a single-qubit quantum channel such that $\operatorname{tr}(\mathcal{E}\circ \operatorname{diag}) \neq 1$. The sample complexity $N_{\text{tot}}$ to estimate the linear target functions $\{\operatorname{tr}(P_i\rho)\}_{i=1}^M$ of an $n$-qubit state $\rho$ within error $\varepsilon$ and failure probability $\delta$ when unitary ensemble $\mathcal{U}$ is subject to quantum channel $\mathcal{E}^{\otimes n}$ is*

$$
N_{\text{tot}} \leq \frac{68\log(2M/\delta)}{\varepsilon^2} \max_{1\leq i\leq M} \Big(\frac{1}{3f(\mathcal{E})^2}\Big)^{\text{wt}(P_i)},
$$

*where $f(\mathcal{E}) = \frac{1}{3}(\operatorname{tr}(\mathcal{E}\circ\operatorname{diag}) - 1)$ and $\text{wt}(P) = |\{i : P_i \neq \mathbb{I}\}|$.*

*Remark* 5.6. We also studied the shadow seminorm of a general $k$-local observable in the presence of a general product channel. However, we could not derive a simple expression. In Appendix D, we derive a bound on the shadow seminorm of a $k$-local observable when the error channel is depolarizing noise (rather than a general product channel).

## 5.4 Examples

We apply the results of this section to derive expressions for the shadow channel, inverse shadow channel, and the classical shadow when the unitary ensemble is the product Clifford ensemble and we fix the quantum channel. Specifically, we study the identity channel, depolarizing channel, and amplitude damping channel. For each quantum channel, we also bound the shadow seminorm for $k$-local Pauli observables, which imply sample complexity bounds for the classical shadows protocol with the product Clifford ensemble in the presence of noise.

### 5.4.1 Noiseless Channel

We start with the noiseless case to show that our results can be used to recover Eq. (S44) through (S50) in [11]. Recall from Section 4.4.1, we show $f(\mathbb{I}) = 1/(2^n+1)$. Hence, for the local case ($n = 1$), $f(\mathbb{I}) = 1/3$. It follows from Claim 5.1 and Claim 5.2, that

$$\mathcal{M}_{\mathcal{C}_1^{\otimes n},\mathbb{I}} = \mathcal{D}_{1,1/3}^{\otimes n}, \qquad \mathcal{M}^{-1}_{\mathcal{C}_1^{\otimes n},\mathbb{I}} = \mathcal{D}_{1,3}^{\otimes n}, \qquad \text{and} \qquad \hat{\rho}(\mathcal{C}_1^{\otimes n}, \mathbb{I}, \hat{U}, \hat{b}) = \bigotimes_{i=1}^{k} (3\hat{U}|\hat{b_i}\rangle\langle\hat{b_i}|\hat{U} - \mathbb{I}).$$

Similarly, by using $f(\mathbb{I}) = 1/3$, the shadow seminorm of a $k$-local Pauli operator can be computed from Proposition 5.4.

$$\|P\|^2_{\text{shadow},\mathcal{C}_1^{\otimes n},\mathbb{I}} = 3^{\text{wt}(P)}.$$

Finally, the sample complexity in this setting follows from Corollary 5.5.

$$N_{\text{tot}} \leq \frac{68\log(2M/\delta)}{\varepsilon^2} \max_{1\leq i\leq M} 3^{\text{wt}(P_i)}.$$

### 5.4.2 Depolarizing Channel

Now we study the case where the quantum channel is $\mathcal{D}_{1,f}^{\otimes n}$ (see Definition 2.2). In Section 4.4.2, we showed that $f(\mathcal{D}_{n,f}) = f/(2^n+1)$, and so, for $n = 1$, we get $f(\mathcal{D}_{1,f}) = f/3$. By applying Claim 5.1 and Claim 5.2, we get the following expressions for the shadow channel, inverse shadow channel, and classical shadow.

$$\mathcal{M}_{\mathcal{C}_1^{\otimes n},\mathcal{D}_{1,f}^{\otimes n}} = \mathcal{D}_{1,f/3}^{\otimes n}, \;\; \mathcal{M}^{-1}_{\mathcal{C}_1^{\otimes n},\mathcal{D}_{1,f}^{\otimes n}} = \mathcal{D}_{1,3/f}^{\otimes n}, \;\; \text{and}$$

$$\hat{\rho}(\mathcal{C}_1^{\otimes n}, \mathcal{D}_{1,f}^{\otimes n}, \hat{U}, \hat{b}) = \bigotimes_{i=1}^{k} \left(\frac{3}{f}\hat{U}|\hat{b_i}\rangle\langle\hat{b_i}|\hat{U} - \left(\frac{1}{2} - \frac{3}{2f}\right)\mathbb{I}\right).$$

Applying Proposition 5.4 with $f(\mathcal{D}_{1,f}) = f/3$, we get

$$\|P\|^2_{\text{shadow},\mathcal{C}_1^{\otimes n},\mathcal{D}_{1,f}^{\otimes n}} = \left(\frac{3}{f^2}\right)^{\text{wt}(P)}.$$

The sample complexity in this setting follows from Corollary 5.5.

$$N_{\text{tot}} \leq \frac{68\log(2M/\delta)}{\varepsilon^2} \max_{1\leq i\leq M} \left(\frac{3}{f^2}\right)^{\text{wt}(P)}.$$

### 5.4.3 Amplitude Damping Channel

The last example we consider is the case where the quantum channel is a product of local amplitude damping channels, denoted by $\text{AD}_{1,p}^{\otimes n}$ (see Definition 2.3). In Section 4.4.3, we show that $f(\text{AD}_{n,p}) = \frac{(1+p)^n-1}{2^{2n}-1}$. For $n = 1$, $f(\text{AD}_{1,p}) = p/3$. We get expressions for the shadow channel and inverse shadow channel by applying Claim 5.1.

$$\mathcal{M}_{\mathcal{C}_1^{\otimes n},\text{AD}_{1,p}^{\otimes n}} = \mathcal{D}_{1,p/3}^{\otimes n} \qquad \text{and} \qquad \mathcal{M}^{-1}_{\mathcal{C}_1^{\otimes n},\text{AD}_{1,p}^{\otimes n}} = \mathcal{D}_{1,3/p}^{\otimes n}.$$

Similarly, we apply Claim 5.2 to get an expression for the classical shadow.

$$\hat{\rho}(\mathcal{C}_1^{\otimes n}, \mathrm{AD}_{1,p}^{\otimes n}, \hat{U}, \hat{b}) = \bigotimes_{i=1}^{k} \left( \frac{3}{p} \hat{U} |\hat{b}_i\rangle\langle\hat{b}_i| \hat{U} - \left( \frac{1}{2} - \frac{3}{2p} \right) \mathbb{I} \right).$$

Applying Proposition 5.4 with $f(\mathrm{AD}_{1,p}) = p/3$, we get

$$\|P\|^2_{\mathrm{shadow},\mathcal{C}_1^{\otimes n},\mathcal{D}_{1,f}^{\otimes n}} = \left(\frac{3}{p^2}\right)^{\mathrm{wt}(P)}.$$

The sample complexity in this setting follows from Corollary 5.5.

$$N_{\mathrm{tot}} \leq \frac{68 \log(2M/\delta)}{\varepsilon^2} \max_{1 \leq i \leq M} \left(\frac{3}{p^2}\right)^{\mathrm{wt}(P)}.$$

## 6 Concluding Remarks and Open Problems

In this paper, we generalized the Huang-Kueng-Preskill classical shadows protocol [11] to take into account the effects of noise. We studied scenarios in which the quantum computer implementing the classical shadows protocol is subject to various noise channels. The noise models we considered include depolarizing noise, dephasing noise and amplitude damping noise.

For each of these noise models, we derived upper bounds for the number of samples needed to achieve expecatation value estimates with a given accuracy. These upper bounds are specified in terms of a shadow seminorm that we introduce in this paper. The shadow seminorm generalizes the shadow norm used to bound the sample complexity in the noiseless classical shadows protocol [11]. By modifying the classical post-processing step of the noiseless protocol, we introduced a new estimator that remains unbiased in the presence of noise. A high-level takeaway of our work is that the classical shadows protocol is still efficient for certain estimation tasks, even in the presence of noise.

We conclude by listing a few open questions and future directions that could build on this work.

1. Comparison of our work with [18]. What if the true noise channel is given by $\mathcal{E}$, but the user thinks that the noise channel is given by $\mathcal{F} \neq \mathcal{E}$? This will result in the application of the inverse shadow channel $\mathcal{M}^{-1}_{\mathcal{U},\mathcal{F}}$ instead of $\mathcal{M}^{-1}_{\mathcal{U},\mathcal{E}}$, which will likely lead to a classical shadow that is not an unbiased estimator of $\rho$. We leave it as an open problem to analyze this setting and give bounds on the bias of the estimator. Furthermore, how does our work compare with the approach given in [18]? If $\mathcal{F}$ and $\mathcal{E}$ are "close" enough, is our approach preferred to [18]?

2. Scope and limitations of our noise model. As noted in Section 1.1.2, the assumption on our noise model is sometimes referred to as the *GTM noise assumption*[8], which allows

[8] For the GTM noise assumption, there is some freedom involved in whether the Markovian noise acts before or after the perfect application of the unitary operation. Indeed, some references (like [19, 20]) have chosen to put the noise before the unitary and others (like [18] and this manuscript) have chosen to put the noise after the unitary. More generally, one could consider the case where known noise channels act both before and after the unitary; in Appendix E, we discuss this case and show that this leads to only a minor modification of the quantities involved in Algorithm 2.

the noisy channel $\widetilde{\mathcal{U}}$ to be written as $\widetilde{\mathcal{U}} = \mathcal{N} \circ \mathcal{U}$, where $\mathcal{U} = U(\cdot)U^\dagger$ is the ideal unitary channel and $\mathcal{N}$ is a quantum channel that is independent of both $U$ and of the physical time at which the computation is performed. While this assumption is common in the literature (see [18–20] and references therein), it is likely too simplistic to represent noise on real devices [109]. If one can give empirical evidence that our algorithm achieves higher accuracy than the original classical shadows protocol, then that would serve as evidence that the GTM assumption is not too simplistic for classical shadows. However, it is still possible that a more realistic noise model could lead to an algorithm that produces higher-accuracy estimates. We leave it as an open problem to analyze classical shadows with more realistic noise models, e.g., noise that is gate-dependent and/or non-Markovian.

3. Experimental demonstration of the classical shadows protocol, where our work is used to improve the accuracy of the estimation. Specifically, this demonstration would involve running experiments for which the estimation accuracy is improved by inverting the noisy shadow channel (rather than the noiseless shadow channel). This would build on some experimental work on classical shadows that have been performed recently, for example, [52, 66, 79].

4. Invertibility of the noisy shadow channel. Are there nice and simple necessary and sufficient conditions for invertibility of the noisy shadow channel? This would generalize the result stated in Section 1.1.1 that a sufficient condition for the noiseless shadow channel to be invertible is that the unitary ensemble is tomographically complete; and would generalize Claim 4.2, which states that if the unitary ensemble is a 2-design and the noise channel is denoted by $\mathcal{E}$, then $\mathrm{tr}(\mathcal{E} \circ \mathrm{diag}) \neq 1$ if and only if the shadow channel is invertible.

5. Comparison of classical shadows with competing methods for estimating properties of quantum states on noisy quantum devices. When should one use classical shadows over other methods? This can be investigated theoretically, where one establishes theoretical performance guarantees for these methods; or numerically or experimentally, where one performs empirical comparisons between the performance of different methods.

   Examples of competing methods include those that we mentioned in Section 1.2.1 and Section 1.2.2. We note here that there has been some recent work along this direction. For example, recent work by Hadfield et al. have compared a non-uniform version of classical shadows (called locally-biased classical shadows) with competing methods like grouping and $\ell^1$-sampling, and have shown that it outperforms these other methods for the task of estimating expectation values of molecular Hamiltonians [50]. An important next step would be to investigate if these advantages continue to hold in the presence of noise.

## Acknowledgements

We thank Peter D. Johnson for useful discussions. We thank Juan Carrasquilla and the anonymous reviewers for their valuable feedback on this manuscript. DEK acknowledges funding support from the National Research Foundation, Singapore, through Grant NRF2021-QEP2-02-P03.

# A Deferred Proofs

## A.1 Proof of Lemma 2.6

To prove Lemma 2.6, we first introduce some notation. For a qudit linear superoperator $\Lambda : \mathcal{L}(\mathbb{C}^d) \to \mathcal{L}(\mathbb{C}^d)$ and $x \in [d]$, define

$$t_{\Lambda,x} \stackrel{\text{def}}{=} \operatorname{tr}(\Lambda(|x\rangle\langle x|)), \tag{93}$$

$$\Lambda_{xxxx} \stackrel{\text{def}}{=} \langle x|\Lambda(|x\rangle\langle x|)|x\rangle. \tag{94}$$

These functions satisfy the following properties:

**Claim A.1.** *Consider the unary operator $()^\ddagger$ defined in Eq. (47). Then,*

*1.*

$$\sum_{x\in[d]} t_{\Lambda,x} = \sum_{x\in[d]} t_{\Lambda^\ddagger,x} = \operatorname{tr}(\Lambda(I)). \tag{95}$$

*If $\Lambda$ is trace-preserving or unital, then*

$$\sum_{x\in[d]} t_{\Lambda,x} = d. \tag{96}$$

*2.*

$$(\Lambda^\ddagger)_{xxxx} = \Lambda_{xxxx}. \tag{97}$$

*3.*

$$\sum_{x\in[d]} \Lambda_{xxxx} = \sum_{x\in[d]} (\Lambda^\ddagger)_{xxxx} = \operatorname{tr}(\Lambda \circ \operatorname{diag}). \tag{98}$$

*Proof.* By straightforward calculation. □

We also need to evaluate the Haar integral in Eq. (20) when $t = 3$. To evaluate the Haar integral, we first introduce some notation. Let $S_3 = \{1, (12), (13), (23), (123), (132)\}$ denote the symmetric group on three elements (where we have written its elements in cycle notation). For each $\pi \in S_3$, define the permutation operator $W_\pi \in \mathbb{U}((\mathbb{C}^d)^{\otimes n})$ to be the unique linear operator satisfying

$$W_\pi(x_1 \otimes x_2 \otimes x_3) = x_{\pi^{-1}(1)} \otimes x_{\pi^{-1}(2)} \otimes x_{\pi^{-1}(3)} \tag{99}$$

for all $x_1, x_2, x_3 \in \mathbb{C}^3$. Equivalently,

$$W_\pi = \sum_{x\in\mathbb{Z}_d^n} |x\rangle\langle\pi(x)|, \tag{100}$$

where $|\pi(x)\rangle = \left|x_{\pi(1)}, \ldots, x_{\pi(n)}\right\rangle$.

Following [110], define the following linear combinations of permutation operators:

$$R_+ = \frac{1}{6}\sum_{\pi\in S_3} W_\pi = \frac{1}{6}(I + W_{12} + W_{13} + W_{23} + W_{123} + W_{132}), \tag{101}$$

$$R_- = \frac{1}{6}\sum_{\pi\in S_3} \operatorname{sgn}(\pi) W_\pi = \frac{1}{6}(I - W_{12} - W_{13} - W_{23} + W_{123} + W_{132}), \tag{102}$$

$$R_0 = \frac{1}{3}(2I - W_{123} - W_{132}), \tag{103}$$

$$R_1 = \frac{1}{3}(2W_{23} - W_{13} - W_{12}), \tag{104}$$

$$R_2 = \frac{1}{\sqrt{3}}(W_{12} - W_{13}), \tag{105}$$

$$R_3 = \frac{\mathrm{i}}{\sqrt{3}}(W_{123} - W_{132}), \tag{106}$$

where we have dropped the parentheses in the notation for the permutation operators: $W_{12} = W_{(12)}, W_{13} = W_{(13)}$, etc.

When $t = 3$, the Haar integral in Eq. (20) may be expressed in terms of the operators $R_i$ as follows [111, Eq. (A3)]:

$$T_3^{(d)}(A) = \frac{6\operatorname{tr}(R_+A)}{d(d+1)(d+2)}R_+ + \frac{6\operatorname{tr}(R_-A)}{d(d-1)(d-2)}R_- + \frac{3}{2d(d^2-1)}\sum_{i=0}^{3}\operatorname{tr}(R_iA)R_i. \tag{107}$$

We now state and prove the following identities.

**Lemma A.2.** *Let $d \in \mathbb{Z}^+$ and $x \in \mathbb{Z}_d$. Let $\Lambda : \mathcal{L}(\mathbb{C}^d) \to \mathcal{L}(\mathbb{C}^d)$ be a linear superoperator, and let $A, B, \Gamma \in \mathcal{L}(\mathbb{C}^d)$ be linear operators. Then,*

$$T_3^{(d)}(|zxx\rangle\langle yxx|) = \frac{2}{d(d+1)(d+2)}(\delta_{yz} + 2\delta_{xyz})R_+ + \frac{1}{d(d+1)(d-1)}(\delta_{yz} - \delta_{xyz})(R_0 + R_1). \tag{108}$$

$$\begin{aligned} T_3^{(d)}(\Gamma \otimes |xx\rangle\langle xx|) = {} & \frac{2}{d(d+1)(d+2)}(\operatorname{tr}\Gamma + 2\langle x|\Gamma|x\rangle)R_+ \\ & + \frac{1}{d(d+1)(d-1)}(\operatorname{tr}\Gamma - \langle x|\Gamma|x\rangle)(R_0 + R_1). \end{aligned} \tag{109}$$

$$\begin{aligned} & \operatorname{tr}_{23}\{T_3^{(d)}(\Lambda(|x\rangle\langle x|) \otimes |xx\rangle\langle xx|)(I \otimes B \otimes C)\} \\ & \quad = \frac{1}{(d-1)d(d+1)(d+2)}\Big\{ ((1+d)t_{\Lambda,x} - 2\Lambda_{xxxx})\,[\operatorname{tr}(BC) + \operatorname{tr}(B)\operatorname{tr}(C)] \\ & \qquad + (d\Lambda_{xxxx} - t_{\Lambda,x}\,[B\operatorname{tr}(C) + C\operatorname{tr}(B) + BC + CB]\Big\}. \end{aligned} \tag{110}$$

*Proof.*

- To prove Eq. (108), let $A = |zxx\rangle\langle yxx|$. Using the definitions of $R_i$ from Eqs. (101)–(106), we can calculate that

$$\operatorname{tr}(R_+A) = \frac{1}{3}(\delta_{yz} + 2\delta_{xyz}), \tag{111}$$

$$\operatorname{tr}(R_-A) = \operatorname{tr}(R_2A) = \operatorname{tr}(R_3A) = 0, \tag{112}$$

$$\operatorname{tr}(R_0A) = \operatorname{tr}(R_1A) = \frac{2}{3}(\delta_{yz} - \delta_{xyz}). \tag{113}$$

Substituting these into Eq. (107) gives Eq. (108).

- To prove Eq. (109), we decompose

$$\Gamma = \sum_{yz} \gamma_{yz} |z\rangle\langle y|. \tag{114}$$

Then, by linearity,

$$T_3^{(d)}(\Gamma \otimes |xx\rangle\langle xx|) = \sum_{yz} \gamma_{yz} T_3^{(d)}(|zxx\rangle\langle yxx|). \tag{115}$$

Eq. (109) follows from Eq. (108) and the facts that

$$\sum_{yz} \gamma_{yz}\delta_{yz} = \operatorname{tr}\Gamma, \tag{116}$$

$$\sum_{yz} \gamma_{yz}\delta_{xyz} = \gamma_{xx} = \langle x|\Gamma|x\rangle. \tag{117}$$

- To prove Eq. (110), set $\Gamma = \Lambda(|x\rangle\langle x|)$. Then, $\operatorname{tr}\Gamma = t_{\Lambda,x}$ and $\langle x|\Gamma|x\rangle = \Lambda_{xxxx}$. Using Eq. (109),

$$\begin{aligned} T_3^{(d)}(\Lambda(|x\rangle\langle x|) \otimes |xx\rangle\langle xx|) = & \frac{2}{d(d+1)(d+2)}(t_{\Lambda,x} + 2\Lambda_{xxxx})R_+ \\ & + \frac{1}{d(d+1)(d-1)}(t_{\Lambda,x} - \Lambda_{xxxx})(R_0 + R_1). \end{aligned} \tag{118}$$

Substituting this into the left-hand-side of Eq. (110) gives

$$\text{LHS} = \frac{2(t_{\Lambda,x} + 2\Lambda_{xxxx})}{d(d+1)(d+2)} \underbrace{\operatorname{tr}_{23}[R_+(I \otimes B \otimes C)]}_{1} + \frac{t_{\Lambda,x} - \Lambda_{xxxx}}{d(d+1)(d-1)} \underbrace{\operatorname{tr}_{23}[(R_0 + R_1)(I \otimes B \otimes C)]}_{2}. \tag{119}$$

Let

$$\xi_\pi = \operatorname{tr}_{23}(W_\pi(I \otimes B \otimes C)). \tag{120}$$

Expanding $R_+, R_0$ and $R_1$, we obtain

$$\begin{aligned} 1 &= \frac{1}{6}\sum_{\pi\in S_3} \xi_\pi \\ &= \tfrac{1}{6}(\operatorname{tr}(B)\operatorname{tr}(C) + \operatorname{tr}(BC) + B\operatorname{tr}(C) + C\operatorname{tr}(B) + BC + CB), \end{aligned} \tag{121}$$

$$2 = \operatorname{tr}_{23}\{[\tfrac{1}{3}(2I - W_{123} - W_{132}) + \tfrac{1}{3}(2W_{23} - W_{13} - W_{12})](I \otimes B \otimes C)\}$$

$$= \frac{1}{3}(2\xi_1 - \xi_{(12)} - \xi_{(13)} + 2\xi_{(23)} - \xi_{(123)} - \xi_{132})$$
$$= \tfrac{1}{3}(2\operatorname{tr}(B)\operatorname{tr}(C) + 2\operatorname{tr}(BC) - B\operatorname{tr}(C) - C\operatorname{tr}(B) - BC - CB), \tag{122}$$

where we used the following identities

$$\xi_1 = \operatorname{tr}(B)\operatorname{tr}(C) \tag{123}$$
$$\xi_{(12)} = B\operatorname{tr}(C) \tag{124}$$
$$\xi_{(13)} = C\operatorname{tr}(B) \tag{125}$$
$$\xi_{(23)} = \operatorname{tr}(BC) \tag{126}$$
$$\xi_{(123)} = CB \tag{127}$$
$$\xi_{(132)} = BC. \tag{128}$$

Substituting Eq. (121) and Eq. (122) into Eq. (119) and rearranging terms gives Eq. (110).

□

We are now ready to prove Lemma 2.6. Let $\mathcal{U}$ be a qudit 3-design and $\mathcal{E} : \mathcal{L}(\mathbb{C}^d) \to \mathcal{L}(\mathbb{C}^d)$ be a linear superoperator. Let $b \in [d]$ and let $A, B, C \in \mathcal{L}(\mathbb{C}^d)$ be linear operators. First, we consider the function

$$\begin{aligned}
\Xi_{\mathcal{E}}(b) :&= \mathop{\mathbb{E}}_{U\sim\mathcal{U}} U^\dagger \mathcal{E}(|b\rangle\langle b|)U \langle b|UBU^\dagger|b\rangle\langle b|UCU^\dagger|b\rangle \\
&= \mathop{\mathbb{E}}_{U\sim\mathcal{U}} U^\dagger \mathcal{E}(|b\rangle\langle b|)U \langle bb|(U\otimes U)(B\otimes C)(U^\dagger\otimes U^\dagger)|bb\rangle \\
&= \operatorname{tr}_{23} \mathop{\mathbb{E}}_{U\sim\mathcal{U}} \left\{U^\dagger \mathcal{E}(|b\rangle\langle b|)U \otimes (U^\dagger\otimes U^\dagger)|bb\rangle\langle bb|(U\otimes U)(B\otimes C)\right\} \\
&= \operatorname{tr}_{23}\left[\mathop{\mathbb{E}}_{U\in\mathcal{U}} (U^\dagger\otimes U^\dagger\otimes U^\dagger)\mathcal{E}(|b\rangle\langle b|)\otimes|bb\rangle\langle bb|(U\otimes U\otimes U)(I\otimes B\otimes C)\right] \\
&= \operatorname{tr}_{23}\left[\mathop{\mathbb{E}}_{U\in\mathcal{U}^\dagger} (U\otimes U\otimes U)\mathcal{E}(|b\rangle\langle b|)\otimes|bb\rangle\langle bb|(U^\dagger\otimes U^\dagger\otimes U^\dagger)(I\otimes B\otimes C)\right] \\
&= \operatorname{tr}_{23}\left[T_3^{(d)}(\mathcal{E}(|b\rangle\langle b|)\otimes|bb\rangle\langle bb|)(I\otimes B\otimes C)\right] \\
&= \frac{1}{(d-1)d(d+1)(d+2)}\Big\{ [(1+d)t_{\mathcal{E},b} - 2\mathcal{E}_{bbbb}]\,[\operatorname{tr}(BC)+\operatorname{tr}(B)\operatorname{tr}(C)] \\
&\quad + (d\mathcal{E}_{bbbb} - t_{\mathcal{E},b})\,[B\operatorname{tr}(C) + C\operatorname{tr}(B) + BC + CB]\Big\}
\end{aligned} \tag{129}$$

where the sixth line follows from the assumption that $\mathcal{U}^\dagger = \{U : U^\dagger \in \mathcal{U}\}$ is a 3-design, and the last line follows from Eq. (110).

Summing $\Xi_{\mathcal{E}}(b)$ over all $b$, we obtain

$$\begin{aligned}
\sum_{b\in[d]} \Xi_{\mathcal{E}}(b) &= \sum_{b\in[d]} \Xi_{\mathcal{E}^\ddagger}(b) \\
&= \frac{1}{(d-1)d(d+1)(d+2)}\Big\{ [(1+d)\operatorname{tr}(\mathcal{E}(I)) - 2\operatorname{tr}(\mathcal{E}\circ\mathrm{diag})]\,[\operatorname{tr}(BC)+\operatorname{tr}(B)\operatorname{tr}(C)] \\
&\quad + (d\operatorname{tr}(\mathcal{E}\circ\mathrm{diag}) - \operatorname{tr}(\mathcal{E}(I)))\,[B\operatorname{tr}(C)+C\operatorname{tr}(B)+BC+CB]\Big\},
\end{aligned} \tag{130}$$

where $\mathcal{E}^\ddagger(A) = (\mathcal{E}^*(A^\dagger))^\dagger$. We use the (easy-to-verify) fact that $\mathcal{E}$ is trace-preserving iff $\mathcal{E}^\ddagger$ is unital, which implies that $\operatorname{tr}(\mathcal{E}(I)) = \operatorname{tr}(\mathcal{E}^\ddagger(I))$ and $\operatorname{tr}(\mathcal{E} \circ \operatorname{diag}) = \operatorname{tr}(\mathcal{E}^\ddagger \circ \operatorname{diag})$.

Hence,

$$
\begin{aligned}
&\mathop{\mathbb{E}}_{U\sim\mathcal{U}} \sum_{b\in[d]} \langle b|\mathcal{E}(UAU^\dagger)|b\rangle\langle b|UBU^\dagger|b\rangle\langle b|UCU^\dagger|b\rangle \\
&= \mathop{\mathbb{E}}_{U\sim\mathcal{U}} \sum_{b\in[d]} \operatorname{tr}\{\mathcal{E}(UAU^\dagger)|b\rangle\langle b|\}\langle b|UBU^\dagger|b\rangle\langle b|UCU^\dagger|b\rangle \\
&= \mathop{\mathbb{E}}_{U\sim\mathcal{U}} \sum_{b\in[d]} \operatorname{tr}\{UAU^\dagger\mathcal{E}^\ddagger(|b\rangle\langle b|)\}\langle b|UBU^\dagger|b\rangle\langle b|UCU^\dagger|b\rangle \\
&= \mathop{\mathbb{E}}_{U\sim\mathcal{U}} \sum_{b\in[d]} \operatorname{tr}\{AU^\dagger\mathcal{E}^\ddagger(|b\rangle\langle b|)U\}\langle b|UBU^\dagger|b\rangle\langle b|UCU^\dagger|b\rangle \\
&= \operatorname{tr}\left\{A \sum_{b\in[d]} \mathop{\mathbb{E}}_{U\sim\mathcal{U}} U^\dagger\mathcal{E}^\ddagger(|b\rangle\langle b|)U\langle b|UBU^\dagger|b\rangle\langle b|UCU^\dagger|b\rangle\right\} \\
&= \operatorname{tr}\left\{A \sum_{b\in[d]} \Xi_{\mathcal{E}^\ddagger}(b)\right\} \\
&= \frac{(1+d)\alpha - 2\beta}{(d-1)d(d+1)(d+2)} \left(\operatorname{tr}(A)\operatorname{tr}(BC) + \operatorname{tr}(A)\operatorname{tr}(B)\operatorname{tr}(C)\right) \\
&\quad + \frac{d\beta - \alpha}{(d-1)d(d+1)(d+2)} \left(\operatorname{tr}(AB)\operatorname{tr}(C) + \operatorname{tr}(AC)\operatorname{tr}(B) + \operatorname{tr}(ABC) + \operatorname{tr}(ACB)\right),
\end{aligned}
\tag{131}
$$

where $\alpha = \operatorname{tr}(\mathcal{E}(I))$ and $\beta = \operatorname{tr}(\mathcal{E} \circ \operatorname{diag})$, and where the last line follows from applying Eq. (130).

## A.2 Proof of Lemma 3.6

Below we give the proof of Lemma 3.6.

*Proof.* By definition, the variance of $\hat{o}$ is

$$\mathop{\operatorname{Var}}_{\substack{U\sim\mathcal{U}\\ b\sim P_b}}[\hat{o}] = \mathop{\mathbb{E}}_{\substack{U\sim\mathcal{U}\\ b\sim P_b}}\Big[\big(\hat{o} - \mathop{\mathbb{E}}_{\substack{U\sim\mathcal{U}\\ b\sim P_b}}[\hat{o}]\big)^2\Big].$$

Let $O_o = O - \operatorname{tr}(O)\frac{\mathbb{I}}{2^n}$ be the traceless part of $O$. Because $\operatorname{tr}(\rho) = \operatorname{tr}(\hat{\rho}) = 1$, it follows that $\hat{o} - \mathbb{E}[\hat{o}] = \operatorname{tr}(O_o\hat{\rho}) = \operatorname{tr}(O_o\rho)$. Therefore, the variance depends only on $O_o$:

$$\mathop{\operatorname{Var}}_{\substack{U\sim\mathcal{U}\\ b\sim P_b}}[\hat{o}] = \mathop{\mathbb{E}}_{\substack{U\sim\mathcal{U}\\ b\sim P_b}}\Big[\big(\operatorname{tr}(O_o\hat{\rho}) - \operatorname{tr}(O_o\rho)\big)^2\Big]. \tag{132}$$

Simplifying further, we get

$$\mathop{\operatorname{Var}}_{\substack{U\sim\mathcal{U}\\ b\sim P_b}}[\hat{o}] = \mathop{\mathbb{E}}_{\substack{U\sim\mathcal{U}\\ b\sim P_b}}\Big[\big(\operatorname{tr}(O_o\hat{\rho}) - \operatorname{tr}(O_o\rho)\big)^2\Big]$$

$$
\begin{aligned}
&= \mathop{\mathbb{E}}_{\substack{U\sim\mathcal{U}\\ b\sim P_b}}\Big[\operatorname{tr}(O_o\hat{\rho})^2 + \operatorname{tr}(O_o\rho)^2 - 2\operatorname{tr}(O_o\rho)\operatorname{tr}(O_o\hat{\rho})\Big] \\
&= \mathop{\mathbb{E}}_{\substack{U\sim\mathcal{U}\\ b\sim P_b}}\Big[\big(\operatorname{tr}(O_o\hat{\rho})^2\Big] + \operatorname{tr}(O_o\rho)^2 - 2\operatorname{tr}(O_o\rho)\mathop{\mathbb{E}}_{\substack{U\sim\mathcal{U}\\ b\sim P_b}}\Big[\operatorname{tr}(O_o\hat{\rho})\Big] \\
&= \mathop{\mathbb{E}}_{\substack{U\sim\mathcal{U}\\ b\sim P_b}}\Big[\operatorname{tr}(O_o\hat{\rho})^2\Big] - \operatorname{tr}(O_o\rho)^2,
\end{aligned} \tag{133}
$$

where

$$
\begin{aligned}
\operatorname{tr}(O_o\hat{\rho}) &= \operatorname{tr}\left(O_o\mathcal{M}^{-1}_{\mathcal{U},\mathcal{E}}(U^\dagger|b\rangle\langle b|U)\right) \\
&= \operatorname{tr}(\mathcal{M}^{-1,\dagger}_{\mathcal{U},\mathcal{E}}(O_o)U^\dagger|b\rangle\langle b|U) \\
&= \langle b|U\mathcal{M}^{-1,\dagger}_{\mathcal{U},\mathcal{E}}(O_o)U^\dagger|b\rangle.
\end{aligned} \tag{134}
$$

To get *a priori* bounds on the variance, we must remove the dependence on the input state $\rho$, which we do by maximizing over all quantum states.

$$
\begin{aligned}
\mathop{\mathrm{Var}}_{\substack{U\sim\mathcal{U}\\ b\sim P_b}}[\hat{o}] &= \mathop{\mathbb{E}}_{\substack{U\sim\mathcal{U}\\ b\sim P_b}}\Big[\langle b|U\mathcal{M}^{-1,\dagger}_{\mathcal{U},\mathcal{E}}(O_o)U^\dagger|b\rangle^2\Big] - \operatorname{tr}(O_o\rho)^2 \\
&= \mathop{\mathbb{E}}_{U\sim\mathcal{U}}\sum_{b\in\{0,1\}^n} P_b(b;U,\mathcal{E},\rho)\langle b|U\mathcal{M}^{-1,\dagger}_{\mathcal{U},\mathcal{E}}(O_o)U^\dagger|b\rangle^2 - \operatorname{tr}(O_o\rho)^2 \\
&= \mathop{\mathbb{E}}_{U\sim\mathcal{U}}\sum_{b\in\{0,1\}^n} \langle b|\mathcal{E}(U\rho U^\dagger)|b\rangle\langle b|U\mathcal{M}^{-1,\dagger}_{\mathcal{U},\mathcal{E}}(O_o)U^\dagger|b\rangle^2 - \operatorname{tr}(O_o\rho)^2 \\
&\leq \max_{\sigma\in\mathbb{D}_{2^n}}\mathop{\mathbb{E}}_{U\sim\mathcal{U}}\sum_{b\in\{0,1\}^n} \langle b|\mathcal{E}(U\sigma U^\dagger)|b\rangle\langle b|U\mathcal{M}^{-1,\dagger}_{\mathcal{U},\mathcal{E}}(O_o)U^\dagger|b\rangle^2 - \operatorname{tr}(O_o\rho)^2 \\
&= \left(\max_{\sigma\in\mathbb{D}_{2^n}}\sqrt{\mathop{\mathbb{E}}_{U\sim\mathcal{U}}\sum_{b\in\{0,1\}^n} \langle b|\mathcal{E}(U\sigma U^\dagger)|b\rangle\langle b|U\mathcal{M}^{-1,\dagger}_{\mathcal{U},\mathcal{E}}(O_o)U^\dagger|b\rangle^2}\right)^2 - \operatorname{tr}(O_o\rho)^2 \\
&= \|O_o\|^2_{\mathrm{shadow},\mathcal{U},\mathcal{E}} - \operatorname{tr}(O_o\rho)^2 \\
&\leq \left\|O - \operatorname{tr}(O)\frac{\mathbb{I}}{2^n}\right\|^2_{\mathrm{shadow},\mathcal{U},\mathcal{E}}.
\end{aligned}
$$

□

### A.3 Proof of Corollary 4.6

*Proof.* We use the fact that if $A \in \mathbb{H}_{2^n}$, then $2^{-n}\operatorname{tr}(A) \leq \|A\|_{\mathrm{sp}} \leq \operatorname{tr}(A)$. First, we prove the lower bound.

$$
\begin{aligned}
\left\|O_o^2\right\|_{\mathrm{sp}} \geq \frac{1}{2^n}\operatorname{tr}(O_o^2) \implies \|O_o\|^2_{\mathrm{shadow},\mathcal{U},\mathcal{E}} &\geq \frac{2^{2n}-1}{(2^n+2)(\beta-1)}\Big(\frac{2^n+2^{2n}-2\beta}{2^n(\beta-1)} + \frac{2}{2^n}\Big)\operatorname{tr}(O_o^2) \\
&= \frac{(2^{2n}-1)(2^n-1)}{2^n(\beta-1)^2}\operatorname{tr}(O_o^2)
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{(2^n-1)^2(2^n+1)}{2^n(\beta-1)^2}\operatorname{tr}(O_o^2) \\
&\geq \frac{(2^n-1)^2}{(\beta-1)^2}\operatorname{tr}(O_o^2). \qquad (135)
\end{aligned}
$$

Now, the first upper bound.

$$
\begin{aligned}
\left\|O_o^2\right\|_{\text{sp}} \leq \operatorname{tr}(O_o^2) \implies \|O_o\|^2_{\text{shadow},\mathcal{U},\mathcal{E}} &\leq \frac{2^{2n}-1}{(2^n+2)(\beta-1)}\Big(\frac{2^n+2^{2n}-2\beta}{2^n(\beta-1)}+2\Big)\operatorname{tr}(O_o^2) \\
&= \frac{(2^n-1)^2(2^n+1)(2\beta+2^n)}{2^n(2^n+2)(\beta-1)^2}\operatorname{tr}(O_o^2) \\
&\leq \frac{(2^n-1)^2(2\beta+2^n)}{2^n(\beta-1)^2}\operatorname{tr}(O_o^2) \\
&\leq \frac{(2^n-1)^2(2\cdot 2^n+2^n)}{2^n(\beta-1)^2}\operatorname{tr}(O_o^2) \\
&= \frac{3(2^n-1)^2}{(\beta-1)^2}\operatorname{tr}(O_o^2). \qquad (136)
\end{aligned}
$$

Finally, the second upper bound.

$$
\begin{aligned}
\|O_o\|^2_{\text{shadow},\mathcal{U},\mathcal{E}} &\leq \frac{3(2^n-1)^2}{(\beta-1)^2}\operatorname{tr}(O_o^2) \\
&= \frac{3(2^n-1)^2}{(\beta-1)^2}\operatorname{tr}\big((O-\frac{1}{2^n}\operatorname{tr}(O)\mathbb{I})^2\big) \\
&= \frac{3(2^n-1)^2}{(\beta-1)^2}\big(\operatorname{tr}(O^2)-\frac{1}{2^n}\operatorname{tr}(O)^2\big) \\
&\leq \frac{3(2^n-1)^2}{(\beta-1)^2}\operatorname{tr}(O^2). \qquad (137)
\end{aligned}
$$

□

# B When is the Shadow Seminorm a Norm?

In this appendix, we prove that the shadow seminorm is indeed a seminorm. In addition, we prove sufficient conditions for the shadow seminorm $\|\cdot\|_{\text{shadow},\mathcal{U},\mathcal{E}}$ to be a norm. The paradigmatic quantum channels studied in this work—namely the depolarizing channel, dephasing channel and amplitude damping channel—satisfy these conditions, and hence yield shadow seminorms which are also norms. We begin by stating the main result of this section.

**Definition B.1.** Let $\Lambda_n$ be the set of $n$-qubit quantum channels $\Theta$ satisfying the following:

$$
\forall b \in \{0,1\}^n,\ \exists\, \sigma \in \mathbb{D}_{2^n} : \langle b|\Theta(\sigma)|b\rangle \neq 0.
$$

**Proposition B.2.** *Let $\mathcal{U}$ be an $n$-qubit unitary ensemble and $\mathcal{E}$ an $n$-qubit quantum channel such that the shadow channel $\mathcal{M}_{\mathcal{U},\mathcal{E}}$ is invertible. Then,*

1. *The shadow seminorm $\|\cdot\|_{\mathrm{shadow},\mathcal{U},\mathcal{E}}$ is a seminorm.*

2. *If $\mathcal{E} \in \Lambda_n$, then the shadow seminorm $\|\cdot\|_{\mathrm{shadow},\mathcal{U},\mathcal{E}}$ is a norm.*

The rest of this appendix is dedicated to proving this statement (which amounts to explicitly verifying that the shadow seminorm satisfies the properties of being a seminorm or norm).

First, we show that the condition above doesn't trivially include all channels (i.e., there are channels that are not in $\Lambda_n$). We also show that all unital channels are in $\Lambda_n$.

**Claim B.3.**

1. *There exist quantum channels which are not in $\Lambda_n$.*

2. *If a quantum channel $\Theta$ is unital, then $\Theta \in \Lambda_n$.*

*Proof.* To prove (1), choose $\Theta$ to be the following map: $A \mapsto \mathrm{tr}(A)|0_n\rangle\langle 0_n|$ ($|0_n\rangle$ denotes the $n$-qubit state where all qubits are in the state $|0\rangle$). From the following Kraus representation of $\Theta$, we see that $\Theta$ is a quantum channel:

$$\Theta(A) = \mathrm{tr}(A)|0_n\rangle\langle 0_n| = \sum_i \langle i|A|i\rangle|0_n\rangle\langle 0_n| = \sum_i |0_n\rangle\langle i|A|0_n\rangle\langle i|^\dagger.$$

Now take $b = 1_n = 11\ldots1$ (that is, $|b\rangle = |1_n\rangle$ is the $n$-qubit state where all qubits are in the state $|1\rangle$). Then, $\forall\sigma \in \mathbb{D}_{2^n}, \langle b|\Theta(\sigma)|b\rangle = 0$. Thus, $\Theta \notin \Lambda_n$.

To prove (2), assume that the quantum channel $\Theta$ is unital. Let $b \in \{0,1\}^n$ and choose $\sigma$ to be the maximally mixed state. Then,

$$\langle b|\Theta(\sigma)|b\rangle = \frac{1}{2^n}\langle b|\Theta(I)|b\rangle = \frac{1}{2^n} \neq 0.$$

□

*Proof of Proposition B.2.*

1. To show that the shadow seminorm is a seminorm, we shall explicitly verify that it satisfies the triangle inequality and absolute homogeneity.

   First, the triangle inequality.

$$\|S+T\|_{\mathrm{shadow},\mathcal{U},\mathcal{E}} = \max_{\sigma\in\mathbb{D}_{2^n}} \sqrt{\underbrace{\mathop{\mathbb{E}}_{U\sim\mathcal{U}} \sum_{b\in\{0,1\}^n} \langle b|\mathcal{E}(U\sigma U^\dagger|b\rangle\langle b|U\mathcal{M}_{\mathcal{U},\mathcal{E}}^{-1,\dagger}(S+T)U^\dagger|b\rangle^2}_{1}}. \quad (138)$$

   Then,

$$\begin{aligned}
1 &= \mathop{\mathbb{E}}_{U\sim\mathcal{U}} \sum_{b\in\{0,1\}^n} \langle b|\mathcal{E}(U\sigma U^\dagger|b\rangle\langle b|U\mathcal{M}_{\mathcal{U},\mathcal{E}}^{-1,\dagger}(S+T)U^\dagger|b\rangle^2 \\
&= \mathop{\mathbb{E}}_{U\sim\mathcal{U}} \sum_{b\in\{0,1\}^n} \langle b|\mathcal{E}(U\sigma U^\dagger|b\rangle\Big(\langle b|U\mathcal{M}_{\mathcal{U},\mathcal{E}}^{-1,\dagger}(S)U^\dagger|b\rangle^2 + |b\rangle\langle b|U\mathcal{M}_{\mathcal{U},\mathcal{E}}^{-1,\dagger}(T)U^\dagger|b\rangle^2\Big) \\
&= \mathop{\mathbb{E}}_{U\sim\mathcal{U}} \sum_{b\in\{0,1\}^n} (\alpha_{b,U} + \beta_{b,U})^2
\end{aligned}$$

$$= \underset{U\sim\mathcal{U}}{\mathbb{E}} \sum_{b\in\{0,1\}^n} \alpha_{b,U}^2 + \underset{U\sim\mathcal{U}}{\mathbb{E}} \sum_{b\in\{0,1\}^n} \beta_{b,U}^2 + 2 \underset{U\sim\mathcal{U}}{\mathbb{E}} \sum_{b\in\{0,1\}^n} \alpha_{b,U}\beta_{b,U}, \tag{139}$$

where we define

$$\alpha_{b,U} \overset{\text{def}}{=} \sqrt{\langle b|\mathcal{E}(U\sigma U^\dagger)|b\rangle}\langle b|U\mathcal{M}_{\mathcal{U},\mathcal{E}}^{-1,\dagger}(S)U^\dagger|b\rangle,$$

$$\beta_{b,U} \overset{\text{def}}{=} \sqrt{\langle b|\mathcal{E}(U\sigma U^\dagger)|b\rangle}\langle b|U\mathcal{M}_{\mathcal{U},\mathcal{E}}^{-1,\dagger}(T)U^\dagger|b\rangle.$$

Then, by the Cauchy-Schwarz inequality,

$$\begin{aligned}
1 &\leq \underset{U\sim\mathcal{U}}{\mathbb{E}} \sum_{b\in\{0,1\}^n} \alpha_{b,U}^2 + \underset{U\sim\mathcal{U}}{\mathbb{E}} \sum_{b\in\{0,1\}^n} \beta_{b,U}^2 + 2\sqrt{\underset{U\sim\mathcal{U}}{\mathbb{E}} \sum_{b\in\{0,1\}^n} \alpha_{b,U}^2}\sqrt{\underset{U\sim\mathcal{U}}{\mathbb{E}} \sum_{b\in\{0,1\}^n} \beta_{b,U}^2} \\
&= \left(\sqrt{\underset{U\sim\mathcal{U}}{\mathbb{E}} \sum_{b\in\{0,1\}^n} \alpha_{b,U}^2} + \sqrt{\underset{U\sim\mathcal{U}}{\mathbb{E}} \sum_{b\in\{0,1\}^n} \beta_{b,U}^2}\right)^2. \qquad (140)
\end{aligned}$$

Plugging this back into the original expression gives

$$\begin{aligned}
\|S+T\|_{\text{shadow},\mathcal{U},\mathcal{E}} &= \max_{\sigma\in\mathbb{D}_{2^n}} \sqrt{\underset{U\sim\mathcal{U}}{\mathbb{E}} \sum_{b\in\{0,1\}^n} \alpha_{b,U}^2} + \sqrt{\underset{U\sim\mathcal{U}}{\mathbb{E}} \sum_{b\in\{0,1\}^n} \beta_{b,U}^2} \\
&\leq \max_{\sigma\in\mathbb{D}_{2^n}} \sqrt{\underset{U\sim\mathcal{U}}{\mathbb{E}} \sum_{b\in\{0,1\}^n} \alpha_{b,U}^2} + \max_{\sigma\in\mathbb{D}_{2^n}} \sqrt{\underset{U\sim\mathcal{U}}{\mathbb{E}} \sum_{b\in\{0,1\}^n} \beta_{b,U}^2} \\
&= \|S\|_{\text{shadow},\mathcal{U},\mathcal{E}} + \|T\|_{\text{shadow},\mathcal{U},\mathcal{E}}\,. \qquad (141)
\end{aligned}$$

Secondly, observe that absolute homogeneity follows immediately by linearity.

2. We shall verify that if $\mathcal{E} \in \Lambda_n$ is satisfied, then the shadow seminorm is point-separating/positive semi-definite, which implies that it is also a norm.

   Suppose that $\|T\|_{\text{shadow},\mathcal{U},\mathcal{E}} = 0$. Then

$$\begin{aligned}
&\max_{\sigma\in\mathbb{D}_{2^n}} \sqrt{\underset{U\sim\mathcal{U}}{\mathbb{E}} \sum_{b\in\{0,1\}^n} \langle b|\mathcal{E}(U\sigma U^\dagger|b\rangle\langle b|U\mathcal{M}_{\mathcal{U},\mathcal{E}}^{-1,\dagger}(T)U^\dagger|b\rangle^2} = 0 \\
&\implies \langle b|\mathcal{E}(U\sigma U^\dagger|b\rangle\langle b|U\mathcal{M}_{\mathcal{U},\mathcal{E}}^{-1,\dagger}(T)U^\dagger|b\rangle^2 = 0, \quad \forall\sigma\in\mathbb{D}_{2^n}, U\in\mathcal{U}, b\in\{0,1\}^n. \qquad (142)
\end{aligned}$$

   Without loss of generality, choose $U = \mathbb{I}$. Since $\mathcal{E} \in \Lambda_n$, we know that $\forall\sigma \in \mathbb{D}_{2^n}$, there is a $b$ such that $\langle b|\mathcal{E}(\sigma)|b\rangle \neq 0$. Therefore, we can conclude that

$$\langle b|\mathcal{M}_{\mathcal{U},\mathcal{E}}^{-1,\dagger}(T)|b\rangle^2 = 0, \quad \forall b\in\{0,1\}^n \implies \mathcal{M}_{\mathcal{U},\mathcal{E}}^{-1,\dagger}(T) = 0 \implies T = 0,$$

   which completes the proof of our proposition.

□

## C Inconsequential Noise

Recall from Lemma 4.1 that when the unitary ensemble is a 2-design, the shadow channel is a depolarizing channel with depolarizing parameter $f(\mathcal{E}) = \frac{1}{2^{2n}-1}\Big(\operatorname{tr}(\mathcal{E}\circ \operatorname{diag}) - \frac{1}{2^n}\operatorname{tr}(\mathcal{E}(\mathbb{I}))\Big)$. In this setting, we characterize inconsequential noise (i.e., the quantum channels that do not affect the classical shadows protocol).

**Claim C.1.** *Let $\mathcal{U}$ be an $n$-qubit 2-design and let $\mathcal{E}$ be a linear superoperator. $\mathcal{E}$ has no effect on $\mathcal{M}_{\mathcal{U}}$ (i.e., $\mathcal{M}_{\mathcal{U},\mathcal{E}} = \mathcal{M}_{\mathcal{U}}$) if and only if $\operatorname{tr}(\mathcal{E}\circ\operatorname{diag}) = 2^n$. Also, $\mathcal{E}$ has no effect on $\mathcal{M}_{\mathcal{U}}$ (i.e., $\mathcal{M}_{\mathcal{U},\mathcal{E}} = \mathcal{M}_{\mathcal{U}}$) if and only if $\langle b|\mathcal{E}(|b\rangle\langle b|)|b\rangle = 1, \forall\, b\in\{0,1\}^n$.*

*Proof.*

$$\mathcal{M}_{\mathcal{U},\mathcal{E}} = \mathcal{M}_{\mathcal{U}} \iff \mathcal{D}_{n,f(\mathcal{E})} = \mathcal{D}_{n,f(\mathbb{I})} \iff f(\mathcal{E}) = f(\mathbb{I}) \iff \operatorname{tr}(\mathcal{E}\circ\operatorname{diag}) = 2^n. \tag{143}$$

Also,

$$\mathcal{M}_{\mathcal{U},\mathcal{E}} = \mathcal{M}_{\mathcal{U}} \iff \operatorname{tr}(\mathcal{E}\circ\operatorname{diag}) = 2^n \iff \langle b|\mathcal{E}(|b\rangle\langle b|)|b\rangle = 1, \forall\, b\in\{0,1\}^n. \tag{144}$$

□

## D Shadow Seminorm of $k$-Local Observable with Product Clifford Ensemble

In this section we bound the shadow seminorm of a $k$-local observable when the product Clifford ensemble is subject to depolarizing noise (rather than a general quantum channel).

**Proposition D.1.** *Let $O \in \mathbb{H}_{2^n}$ be a $k$-local observable with nontrivial part $\widetilde{O}$. Let $\widetilde{O} = \sum_{\mathbf{p}\in\mathbb{Z}_4^k}\alpha_{\mathbf{p}}P_{\mathbf{p}}$ be the expansion of $\widetilde{O}$ in the Pauli basis. Let $0\le f\le 1$. Then,*

$$\|O\|^2_{\text{shadow},\mathcal{C}_1^{\otimes n},\mathcal{D}_{1,f}^{\otimes n}} = \Bigg\|\sum_{\mathbf{p},\mathbf{q}\in\mathbb{Z}_4^k}\alpha_{\mathbf{p}}\alpha_{\mathbf{q}}\widetilde{\mathcal{F}}(\mathbf{p},\mathbf{q})P_{\mathbf{p}}P_{\mathbf{q}}\Bigg\|_{\text{sp}}, \tag{145}$$

*where*

$$\widetilde{\mathcal{F}}(\mathbf{p},\mathbf{q}) = \prod_{j=1}^{k}\widetilde{f}(\mathbf{p}_j,\mathbf{q}_j), \tag{146}$$

*with*

$$\widetilde{f}(p,q) = \begin{cases} 1/f & \text{if } p=q=0. \\ 1 & \text{if } (p=0)\oplus(q=0). \\ 3/f^2 & \text{if } p=q\neq 0. \\ 0 & \text{otherwise.}\end{cases} \tag{147}$$

*Proof.* Without loss of generality, we write $O = \widetilde{O} \otimes \mathbb{I}$, where $\widetilde{O} = \sum_{\mathbf{p} \in \mathbb{Z}_4^k} \alpha_{\mathbf{p}} P_{\mathbf{p}}$. Then,

$$\begin{aligned}
&\|O\|^2_{\text{shadow},\mathcal{C}_1^{\otimes n},\mathcal{D}_{1,f}^{\otimes n}} \\
&\quad = \left\|\widetilde{O}\right\|^2_{\text{shadow},\mathcal{C}_1^{\otimes k},\mathcal{D}_{1,f}^{\otimes k}} \\
&\quad = \max_{\sigma \in \mathbb{D}_{2^k}} \mathop{\mathbb{E}}_{U \sim \mathcal{C}_1^{\otimes k}} \sum_{b \in \{0,1\}^k} \langle b|\mathcal{D}_{1,f}^{\otimes k}(U\sigma U^\dagger)|b\rangle\langle b|U\mathcal{D}_{1,3/f}^{\otimes k}(\widetilde{O})U^\dagger|b\rangle^2 \\
&\quad = \max_{\sigma \in \mathbb{D}_{2^k}} \mathop{\mathbb{E}}_{U \sim \mathcal{C}_1^{\otimes k}} \sum_{b \in \{0,1\}^k} \langle b|\mathcal{D}_{1,f}^{\otimes k}(U\sigma U^\dagger)|b\rangle\langle b|U\mathcal{D}_{1,3/f}^{\otimes k}\left(\sum_{\mathbf{p} \in \mathbb{Z}_4^k} \alpha_{\mathbf{p}} P_{\mathbf{p}}\right)U^\dagger|b\rangle^2 \\
&\quad = \max_{\sigma \in \mathbb{D}_{2^k}} \sum_{\mathbf{p},\mathbf{q} \in \mathbb{Z}_4^k} \alpha_{\mathbf{p}}\alpha_{\mathbf{q}} \underbrace{\mathop{\mathbb{E}}_{U \sim \mathcal{C}_1^{\otimes k}} \sum_{b \in \{0,1\}^k} \langle b|\mathcal{D}_{1,f}^{\otimes k}(U\sigma U^\dagger)|b\rangle\langle b|U\mathcal{D}_{1,3/f}^{\otimes k}(P_{\mathbf{p}})U^\dagger|b\rangle\langle b|U\mathcal{D}_{1,3/f}^{\otimes k}(P_{\mathbf{q}})U^\dagger|b\rangle}_{1}.
\end{aligned} \tag{148}$$

To evaluate 1, write $\sigma = \sum_{i_1 \ldots i_k} \sigma_{i_1 \ldots i_k} e_{i_1} \otimes \ldots \otimes e_{i_k}$. Then,

$$\begin{aligned}
1 &= \mathop{\mathbb{E}}_{U \sim \mathcal{C}_1^{\otimes k}} \sum_{b \in \{0,1\}^k} \langle b|\mathcal{D}_{1,f}^{\otimes k}(U \sum_{i_1 \ldots i_k} \sigma_{i_1 \ldots i_k} e_{i_1} \otimes \ldots \otimes e_{i_k} U^\dagger)|b\rangle \\
&\qquad \cdot \langle b|U\mathcal{D}_{1,3/f}^{\otimes k}(P_{\mathbf{p}})U^\dagger|b\rangle\langle b|U\mathcal{D}_{1,3/f}^{\otimes k}(P_{\mathbf{q}})U^\dagger|b\rangle \\
&= \sum_{i_1 \ldots i_k} \sigma_{i_1 \ldots i_k} \bigotimes_{j=1}^{k} \mathop{\mathbb{E}}_{U_j \sim \mathcal{C}_1} \sum_{b_j \in \{0,1\}} \langle b_j|\mathcal{D}_{1,f}(U_j e_{i_j} U_j^\dagger)|b_j\rangle \\
&\qquad \cdot \langle b_j|U_j\mathcal{D}_{1,3/f}(P_{\mathbf{p}_j})U_j^\dagger|b_j\rangle\langle b_j|U_j\mathcal{D}_{1,3/f}(P_{\mathbf{q}_j})U_j^\dagger|b_j\rangle \\
&= \sum_{i_1 \ldots i_k} \sigma_{i_1 \ldots i_k} \prod_{j=1}^{k} \mathop{\mathbb{E}}_{U_j \sim \mathcal{C}_1} \sum_{b_j \in \{0,1\}} \langle b_j|(fU_j e_{i_j} U_j^\dagger + (1-f)\operatorname{tr}(U_j e_{i_j} U_j^\dagger)\frac{\mathbb{I}}{2})|b_j\rangle \\
&\qquad \cdot \langle b_j|U_j\mathcal{D}_{1,3/f}(P_{\mathbf{p}_j})U_j^\dagger|b_j\rangle\langle b_j|U_j\mathcal{D}_{1,3/f}(P_{\mathbf{q}_j})U_j^\dagger|b_j\rangle \\
&= \sum_{i_1 \ldots i_k} \sigma_{i_1 \ldots i_k} \prod_{j=1}^{k} \mathop{\mathbb{E}}_{U_j \sim \mathcal{C}_1} \sum_{b_j \in \{0,1\}} \left\{f\langle b_j|U_j e_{i_j} U_j^\dagger|b_j\rangle + \frac{1-f}{2}\operatorname{tr}(e_{i_j})\right\} \\
&\qquad \cdot \langle b_j|U_j\mathcal{D}_{1,3/f}(P_{\mathbf{p}_j})U_j^\dagger|b_j\rangle\langle b_j|U_j\mathcal{D}_{1,3/f}(P_{\mathbf{q}_j})U_j^\dagger|b_j\rangle \\
&= \sum_{i_1 \ldots i_k} \sigma_{i_1 \ldots i_k} \prod_{j=1}^{k} \Bigg\{f \underbrace{\mathop{\mathbb{E}}_{U_j \sim \mathcal{C}_1} \sum_{b_j \in \{0,1\}} \langle b_j|U_j e_{i_j} U_j^\dagger|b_j\rangle\langle b_j|U_j\mathcal{D}_{1,3/f}(P_{\mathbf{p}_j})U_j^\dagger|b_j\rangle\langle b_j|U_j\mathcal{D}_{1,3/f}(P_{\mathbf{q}_j})U_j^\dagger|b_j\rangle}_{2} \\
&\qquad + \frac{1-f}{2}\operatorname{tr}(e_{i_j}) \underbrace{\mathop{\mathbb{E}}_{U_j \sim \mathcal{C}_1} \sum_{b_j \in \{0,1\}} \langle b_j|U_j\mathcal{D}_{1,3/f}(P_{\mathbf{p}_j})U_j^\dagger|b_j\rangle\langle b_j|U_j\mathcal{D}_{1,3/f}(P_{\mathbf{q}_j})U_j^\dagger|b_j\rangle}_{3} \Bigg\}.
\end{aligned} \tag{149}$$

To evaluate $\textcircled{2}$ and $\textcircled{3}$, define $\xi[A](p,q)$ as

$$\xi[A](p,q) \overset{\text{def}}{=} \mathop{\mathbb{E}}_{U\sim\mathcal{C}_1} \sum_{b\in\{0,1\}} \langle b|UAU^\dagger|b\rangle\langle b|U\mathcal{D}_{1,3/f}(P_p)U^\dagger|b\rangle\langle b|U\mathcal{D}_{1,3/f}(P_q)U^\dagger|b\rangle.$$

Hence, $\textcircled{2} = \xi[e_{i_j}](\mathbf{p}_j,\mathbf{q}_j)$ and $\textcircled{3} = \xi[\mathbb{I}](\mathbf{p}_j,\mathbf{q}_j)$. We will now find an expression for $\xi[A](p,q)$.

*Case 1:* $p = q = 0$.

$$\begin{aligned}
\xi[A](0,0) &= \mathop{\mathbb{E}}_{U\sim\mathcal{C}_1} \sum_{b\in\{0,1\}} \langle b|UAU^\dagger|b\rangle\langle b|U\mathcal{D}_{1,3/f}(\mathbb{I})U^\dagger|b\rangle\langle b|U\mathcal{D}_{1,3/f}(\mathbb{I})U^\dagger|b\rangle \\
&= \mathop{\mathbb{E}}_{U\sim\mathcal{C}_1} \sum_{b\in\{0,1\}} \langle b|UAU^\dagger|b\rangle \\
&= \operatorname{tr}(A).
\end{aligned} \tag{150}$$

*Case 2:* $p \neq 0, q = 0$.

$$\begin{aligned}
\xi[A](p,0) &= \mathop{\mathbb{E}}_{U\sim\mathcal{C}_1} \sum_{b\in\{0,1\}} \langle b|UAU^\dagger|b\rangle\langle b|U\mathcal{D}_{1,3/f}(P_p)U^\dagger|b\rangle\langle b|U\mathcal{D}_{1,3/f}(\mathbb{I})U^\dagger|b\rangle \\
&= \frac{3}{f} \mathop{\mathbb{E}}_{U\sim\mathcal{C}_1} \sum_{b\in\{0,1\}} \langle b|UAU^\dagger|b\rangle\langle b|UP_pU^\dagger|b\rangle \\
&= \frac{3}{f} \operatorname{tr}\left\{A \mathop{\mathbb{E}}_{U\sim\mathcal{C}_1} \sum_{b\in\{0,1\}} U^\dagger|b\rangle\langle b|U\langle b|UP_pU^\dagger|b\rangle\right\} \\
&= \frac{3}{f} \operatorname{tr}\left\{A\mathcal{M}_{\mathcal{C}_1,\mathbb{I}}(P_p)\right\} \\
&= \frac{1}{f} \operatorname{tr}(AP_p).
\end{aligned} \tag{151}$$

*Case 3:* $p = 0, q \neq 0$. By symmetry,

$$\xi[A](0,q) = \frac{1}{f} \operatorname{tr}(AP_q). \tag{152}$$

*Case 4:* $p \neq 0, q \neq 0$.

$$\begin{aligned}
\xi[A](p,q) &= \mathop{\mathbb{E}}_{U\sim\mathcal{C}_1} \sum_{b\in\{0,1\}} \langle b|UAU^\dagger|b\rangle\langle b|U\mathcal{D}_{1,3/f}(P_p)U^\dagger|b\rangle\langle b|U\mathcal{D}_{1,3/f}(P_q)U^\dagger|b\rangle \\
&= \frac{9}{f^2} \operatorname{tr}\left\{A \mathop{\mathbb{E}}_{U\sim\mathcal{C}_1} \sum_{b\in\{0,1\}} U^\dagger|b\rangle\langle b|U\langle b|UP_pU^\dagger|b\rangle\langle b|U\mathcal{P}_qU^\dagger|b\rangle\right\} \\
&= \frac{9}{f^2} \operatorname{tr}\left\{A\frac{1}{3}\delta_{pq}\mathbb{I}\right\} \\
&= \frac{3}{f^2}\delta_{pq} \operatorname{tr}(A).
\end{aligned} \tag{153}$$

The third equality follows from Eq. (S36) of [11], which itself follows from Eq. (129) by setting $\mathcal{E} = I$, $d = 2$, and $\operatorname{tr}(B) = \operatorname{tr}(C) = 0$. Combining the four cases gives

$$\xi[A](p,q) = \begin{cases} \operatorname{tr}(A) & \text{if } p=q=0. \\ \frac{1}{f}\operatorname{tr}(AP_p) & \text{if } p\neq 0, q=0. \\ \frac{1}{f}\operatorname{tr}(AP_q) & \text{if } p=0, q\neq 0. \\ \frac{3}{f^2}\operatorname{tr}(A) & \text{if } p=q\neq 0. \\ 0 & \text{if } p\neq q, p\neq 0, q\neq 0. \end{cases} = \zeta(p,q)\operatorname{tr}(AP_pP_q), \tag{154}$$

where

$$\zeta(p,q) = \begin{cases} 1/f & \text{if } p=0 \text{ or } q=0. \\ 3/f^2 & \text{if } p=q\neq 0. \\ 0 & \text{otherwise.} \end{cases} \tag{155}$$

Applying this to 2 and 3 gives

$$2 = \xi[e_{i_j}](\mathbf{p}_j,\mathbf{q}_j) = \zeta(\mathbf{p}_j,\mathbf{q}_j)\operatorname{tr}(e_{i_j}P_{\mathbf{p}_j}P_{\mathbf{q}_j})$$

and

$$3 = \xi[\mathbb{I}](\mathbf{p}_j,\mathbf{q}_j) = \zeta(\mathbf{p}_j,\mathbf{q}_j)\operatorname{tr}(P_{\mathbf{p}_j}P_{\mathbf{q}_j}) = 2\zeta(\mathbf{p}_j,\mathbf{q}_j)\delta_{\mathbf{p}_j\mathbf{q}_j}.$$

Plugging these expressions into 1 gives

$$\begin{aligned} 1 &= \sum_{i_1\ldots i_k}\sigma_{i_1\ldots i_k}\prod_{j=1}^{k}\left\{f\zeta(\mathbf{p}_j,\mathbf{q}_j)\operatorname{tr}(e_{i_j}P_{\mathbf{p}_j}P_{\mathbf{q}_j}) + \frac{1-f}{2}\operatorname{tr}(e_{i_j})2\zeta(\mathbf{p}_j,\mathbf{q}_j)\delta_{\mathbf{p}_j\mathbf{q}_j}\right\} \\ &= \sum_{i_1\ldots i_k}\sigma_{i_1\ldots i_k}\Big(\prod_{j=1}^{k}\zeta(\mathbf{p}_j,\mathbf{q}_j)\Big)\prod_{j=1}^{k}\Big\{\underbrace{f\operatorname{tr}(e_{i_j}P_{\mathbf{p}_j}P_{\mathbf{q}_j}) + (1-f)\operatorname{tr}(e_{i_j})\delta_{\mathbf{p}_j\mathbf{q}_j}}_{4}\Big\}. \end{aligned} \tag{156}$$

When $\mathbf{p}_j\neq\mathbf{q}_j$, $4 = f\operatorname{tr}(e_{i_j}P_{\mathbf{p}_j}P_{\mathbf{q}_j})$. When $\mathbf{p}_j=\mathbf{q}_j$, $4 = f\operatorname{tr}(e_{i_j}P_{\mathbf{p}_j}^2) + (1-f)\operatorname{tr}(e_{i_j}) = \operatorname{tr}(e_{i_j})$. Hence,

$$4 = f^{\mathbb{1}_{\mathbf{p}_j\neq\mathbf{q}_j}}\operatorname{tr}(e_{i_j}P_{\mathbf{p}_j}P_{\mathbf{q}_j}).$$

$$\begin{aligned} 1 &= \sum_{i_1\ldots i_k}\sigma_{i_1\ldots i_k}\Big(\prod_{j=1}^{k}\zeta(\mathbf{p}_j,\mathbf{q}_j)\Big)\prod_{j=1}^{k}\Big\{f^{\mathbb{1}_{\mathbf{p}_j\neq\mathbf{q}_j}}\operatorname{tr}(e_{i_j}P_{\mathbf{p}_j}P_{\mathbf{q}_j})\Big\} \\ &= \prod_{j=1}^{k}\underbrace{f^{\mathbb{1}_{\mathbf{p}_j\neq\mathbf{q}_j}}\zeta(\mathbf{p}_j,\mathbf{q}_j)}_{5}\sum_{i_1\ldots i_k}\sigma_{i_1\ldots i_k}\operatorname{tr}((e_{i_1}\otimes\ldots\otimes e_{i_k})(P_{\mathbf{p}_1}\otimes\ldots\otimes P_{\mathbf{p}_k})(P_{\mathbf{q}_1}\otimes\ldots\otimes P_{\mathbf{q}_k})). \end{aligned} \tag{157}$$

Let $\widetilde{f}(p,q) = f^{\mathbb{1}_{\mathbf{p}_j\neq\mathbf{q}_j}}\zeta(\mathbf{p}_j,\mathbf{q}_j)$. Then,

$$\widetilde{f}(p,q) = \begin{cases} f & \text{if } p\neq q. \\ 1 & \text{if } p=q. \end{cases} \times \begin{cases} 1/f & \text{if } p=0 \text{ or } q=0. \\ 3/f^2 & \text{if } p=q\neq 0. \\ 0 & \text{otherwise.} \end{cases} = \begin{cases} 1/f & \text{if } p=q=0. \\ 1 & \text{if } (p=0)\oplus(q=0). \\ 3/f^2 & \text{if } p=q\neq 0. \\ 0 & \text{otherwise.} \end{cases} \tag{158}$$

and $\textcircled{5} = \widetilde{f}(\mathbf{p}_j, \mathbf{q}_j)$. Let $\widetilde{\mathcal{F}}(\mathbf{p}, \mathbf{q}) = \prod_{j=1}^{k} \widetilde{f}(\mathbf{p}_j, \mathbf{q}_j)$. Then,

$$\begin{aligned}\textcircled{1} &= \widetilde{\mathcal{F}}(\mathbf{p}, \mathbf{q}) \sum_{i_1 \dots i_k} \sigma_{i_1 \dots i_k} \operatorname{tr}((e_{i_1} \otimes \dots \otimes e_{i_k}) P_{\mathbf{p}} P_{\mathbf{q}}) \\ &= \widetilde{\mathcal{F}}(\mathbf{p}, \mathbf{q}) \operatorname{tr}(\sigma P_{\mathbf{p}} P_{\mathbf{q}}). \end{aligned} \tag{159}$$

Plugging the expression for $\textcircled{1}$ into Eq. (148) completes the proof.

$$\begin{aligned}\|O\|^2_{\mathrm{shadow}, \mathcal{C}_1^{\otimes n}, \mathcal{D}_{1,f}^{\otimes n}} &= \max_{\sigma \in \mathbb{D}_{2^k}} \sum_{\mathbf{p}, \mathbf{q} \in \mathbb{Z}_4^k} \alpha_{\mathbf{p}} \alpha_{\mathbf{q}} \widetilde{\mathcal{F}}(\mathbf{p}, \mathbf{q}) \operatorname{tr}(\sigma P_{\mathbf{p}} P_{\mathbf{q}}) \\ &= \max_{\sigma \in \mathbb{D}_{2^k}} \operatorname{tr} \left\{ \sigma \sum_{\mathbf{p}, \mathbf{q} \in \mathbb{Z}_4^k} \alpha_{\mathbf{p}} \alpha_{\mathbf{q}} \widetilde{\mathcal{F}}(\mathbf{p}, \mathbf{q}) P_{\mathbf{p}} P_{\mathbf{q}} \right\} \\ &= \left\| \sum_{\mathbf{p}, \mathbf{q} \in \mathbb{Z}_4^k} \alpha_{\mathbf{p}} \alpha_{\mathbf{q}} \widetilde{\mathcal{F}}(\mathbf{p}, \mathbf{q}) P_{\mathbf{p}} P_{\mathbf{q}} \right\|_{\mathrm{sp}}. \end{aligned} \tag{160}$$

□

## E Noisy Input States

Throughout the main text, our assumption has been that the input state $\rho$ is prepared without any errors. But what if the input state given is itself noisy? In this appendix, we consider the case where in addition to the noise described in the noisy measurement primitive of Definition 3.1, the input state $\rho$ is subject to the noise channel $\mathcal{K}$. In other words, instead of the intended transformation $\rho \mapsto U \rho U^\dagger$, the input state transforms as $\rho \mapsto \mathcal{E}(U \mathcal{K}(\rho) U^\dagger)$. This scenario is equivalent to a noise model where the unitary operation $U$ is replaced by one where a noise channel acts both before and after the perfect implementation of $U$ (see Footnote 8).

With this change, Eq. (25) becomes

$$U^\dagger |\hat{b}\rangle\langle\hat{b}| U \quad \text{with probability} \quad P_b(\hat{b}) \stackrel{\text{def}}{=} \langle\hat{b}|\mathcal{E}(U\mathcal{K}(\rho)U^\dagger)|\hat{b}\rangle \quad \text{where} \quad \hat{b} \in \{0,1\}^n. \tag{161}$$

The noisy shadow channel in Eq. (26) is modified to

$$\mathcal{M}_{\mathcal{U},\mathcal{E},\mathcal{K}}(\rho) \stackrel{\text{def}}{=} \mathop{\mathbb{E}}_{U \sim \mathcal{U}} \sum_{b \in \{0,1\}^n} \langle b|\mathcal{E}(U\mathcal{K}(\rho)U^\dagger)|b\rangle U^\dagger |b\rangle\langle b| U = (\mathcal{M}_{\mathcal{U},\mathcal{E}} \circ \mathcal{K})(\rho). \tag{162}$$

and the noisy classical shadow of Eq. (27) becomes

$$\begin{aligned}\hat{\rho} = \hat{\rho}(\mathcal{U}, \mathcal{E}, \mathcal{K}, \hat{U}, \hat{b}) &\stackrel{\text{def}}{=} \mathcal{M}^{-1}_{\mathcal{U},\mathcal{E},\mathcal{K}}(\hat{U}^\dagger |\hat{b}\rangle\langle\hat{b}| \hat{U}) \\ &= (\mathcal{K}^{-1} \circ \mathcal{M}^{-1}_{\mathcal{U},\mathcal{E}})(\hat{U}^\dagger |\hat{b}\rangle\langle\hat{b}| \hat{U}), \end{aligned} \tag{163}$$

where we have assumed that both the shadow channel and the noise channel $\mathcal{K}$ are invertible linear superoperators. As before, we do not assume that the inverses are themselves quantum channels.

Next, the shadow seminorm of Eq. (30) becomes modified to

$$\begin{aligned}\|O\|_{\mathrm{shadow},\mathcal{U},\mathcal{E},\mathcal{K}} &= \max_{\sigma\in\mathbb{D}_{2^n}}\sqrt{\mathbb{E}_{U\sim\mathcal{U}}\sum_{b\in\{0,1\}^n}\langle b|\mathcal{E}(U\sigma U^\dagger)|b\rangle\langle b|U\mathcal{M}^{-1,\dagger}_{\mathcal{U},\mathcal{E},\mathcal{K}}(O)U^\dagger|b\rangle^2} \\ &= \max_{\sigma\in\mathbb{D}_{2^n}}\sqrt{\mathbb{E}_{U\sim\mathcal{U}}\sum_{b\in\{0,1\}^n}\langle b|\mathcal{E}(U\sigma U^\dagger)|b\rangle\langle b|U\mathcal{K}^{-1,\dagger}(\mathcal{M}^{-1,\dagger}_{\mathcal{U},\mathcal{E}}(O))U^\dagger|b\rangle^2}. \end{aligned} \tag{164}$$

Hence, by following the same argument as in the main text, we arrive at Theorem 3.7, but with the shadow seminorm $\|\cdot\|_{\mathrm{shadow},\mathcal{U},\mathcal{E}}$ replaced by $\|\cdot\|_{\mathrm{shadow},\mathcal{U},\mathcal{E},\mathcal{K}}$. Consequently, the only changes needed to adapt Algorithm 2 to this case are

1. In step 2 of Algorithm 2, replace the shadow seminorm with Eq. (164).
2. In step 5 of Algorithm 2, replace $\mathcal{E}(U\rho U^\dagger)$ with $\mathcal{E}(U\mathcal{K}(\rho)U^\dagger)$.
3. In step 8 of Algorithm 2, replace each occurrence of $\mathcal{M}^{-1}_{\mathcal{U},\mathcal{E}}$ with $\mathcal{K}^{-1}\circ\mathcal{M}^{-1}_{\mathcal{U},\mathcal{E}}$.

## References


[1] John Preskill. Quantum Computing in the NISQ era and beyond. *Quantum*, 2:79, 2018. doi:10.22331/q-2018-08-06-79. [pp. 3, 8]

[2] Kishor Bharti, Alba Cervera-Lierta, Thi Ha Kyaw, Tobias Haug, Sumner Alperin-Lea, Abhinav Anand, Matthias Degroote, Hermanni Heimonen, Jakob S. Kottmann, Tim Menke, Wai-Keong Mok, Sukin Sim, Leong-Chuan Kwek, and Alán Aspuru-Guzik. Noisy intermediate-scale quantum algorithms. *Rev. Mod. Phys.*, 94:015004, Feb 2022. doi:10.1103/RevModPhys.94.015004. [pp. 3, 8]

[3] Marco Cerezo, Andrew Arrasmith, Ryan Babbush, Simon C Benjamin, Suguru Endo, Keisuke Fujii, Jarrod R McClean, Kosuke Mitarai, Xiao Yuan, Lukasz Cincio, et al. Variational quantum algorithms. *Nature Reviews Physics*, 3(9):625–644, 2021. doi:10.1038/s42254-021-00348-9. [p. 3]

[4] Alberto Peruzzo, Jarrod McClean, Peter Shadbolt, Man-Hong Yung, Xiao-Qi Zhou, Peter J. Love, Alán Aspuru-Guzik, and Jeremy L. O'Brien. A variational eigenvalue solver on a photonic quantum processor. *Nature communications*, 5:4213, 2014. doi:10.1038/ncomms5213. [p. 3]

[5] Edward Farhi, Jeffrey Goldstone, and Sam Gutmann. A Quantum Approximate Optimization Algorithm. *arXiv preprint arXiv:1411.4028*, 2014. doi:10.48550/arXiv.1411.4028. [p. 3]

[6] Yudong Cao, Jonathan Romero, Jonathan P. Olson, Matthias Degroote, Peter D. Johnson, Mária Kieferová, Ian D. Kivlichan, Tim Menke, Borja Peropadre, Nicolas P.D. Sawaya, et al. Quantum Chemistry in the Age of Quantum Computing. *Chemical reviews*, 119(19):10856–10915, 2019. doi:10.1021/acs.chemrev.8b00803. [p. 3]

[7] Vittorio Giovannetti, Seth Lloyd, and Lorenzo Maccone. Quantum metrology. *Physical review letters*, 96(1):010401, 2006. doi:10.1103/PhysRevLett.96.010401. [p. 3]

[8] Nikolaj Moll, Panagiotis Barkoutsos, Lev S. Bishop, Jerry M. Chow, Andrew Cross, Daniel J. Egger, Stefan Filipp, Andreas Fuhrer, Jay M. Gambetta, Marc Ganzhorn, et al. Quantum optimization using variational algorithms on near-term quantum devices. *Quantum Science and Technology*, 3(3):030503, 2018. https://doi:10.1088/2058-9565/aab822. [p. 3]

[9] Dave Wecker, Matthew B. Hastings, and Matthias Troyer. Progress towards practical quantum variational algorithms. *Physical Review A*, 92(4):042303, 2015. doi:10.1103/PhysRevA.92.042303. [p. 3]

[10] William J. Huggins, Jarrod R. McClean, Nicholas C. Rubin, Zhang Jiang, Nathan Wiebe, K. Birgitta Whaley, and Ryan Babbush. Efficient and noise resilient measurements for quantum chemistry on near-term quantum computers. *npj Quantum Information*, 7(1):1–9, 2021. doi:10.1038/s41534-020-00341-7. [p. 3]

[11] 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, 2020. doi:10.1038/s41567-020-0932-7. [pp. 3, 4, 5, 6, 7, 8, 12, 13, 14, 15, 18, 26, 30, 32, 33, 46]

[12] Jeongwan Haah, Aram Harrow, Zhengfeng Ji, Xiaodi Wu, and Nengkun Yu. Sample-Optimal Tomography of Quantum States. *IEEE Transactions on Information Theory*, 63(9):5628–5641, 2017. doi:10.1109/TIT.2017.2719044. [pp. 3, 7]

[13] Ryan O'Donnell and John Wright. Efficient quantum tomography. In *Proceedings of the forty-eighth annual ACM symposium on Theory of Computing*, pages 899–912, 2016. doi:10.1145/2897518.2897544. [pp. 3, 7]

[14] Scott Aaronson. Shadow Tomography of Quantum States. *SIAM Journal on Computing*, 49(5):STOC18–368, 2019. doi:10.1137/18M120275X. [pp. 3, 8]

[15] Mark R. Jerrum, Leslie G. Valiant, and Vijay V. Vazirani. Random Generation of Combinatorial Structures from a Uniform Distribution. *Theoretical Computer Science*, 43:169–188, 1986. doi:10.1016/0304-3975(86)90174-X. [p. 4]

[16] Huangjun Zhu, Richard Kueng, Markus Grassl, and David Gross. The Clifford group fails gracefully to be a unitary 4-design. *arXiv preprint arXiv:1609.08172*, 2016. doi:10.48550/arXiv.1609.08172. [pp. 5, 12]

[17] Zak Webb. The Clifford group forms a unitary 3-design. *Quantum Information & Computation*, 16(15&16):1379–1400, 2016. doi:10.26421/QIC16.15-16-8. [pp. 5, 12]

[18] Senrui Chen, Wenjun Yu, Pei Zeng, and Steven T. Flammia. Robust Shadow Estimation. *PRX Quantum*, 2:030348, Sep 2021. doi:10.1103/PRXQuantum.2.030348. [pp. 5, 8, 33, 34]

[19] Steven T. Flammia and Joel J. Wallman. Efficient Estimation of Pauli Channels. *ACM Transactions on Quantum Computing*, 1(1):1–32, 2020. doi:10.1145/3408039. [pp. 5, 8, 33, 34]

[20] Senrui Chen, Sisi Zhou, Alireza Seif, and Liang Jiang. Quantum advantages for Pauli channel estimation. *Physical Review A*, 105(3):032435, 2022. doi:10.1103/PhysRevA.105.032435. [pp. 5, 33, 34]

[21] Michael A. Nielsen and Isaac L. Chuang. *Quantum Computation and Quantum Information*. Cambridge University Press, 2010. doi:10.1017/CBO9780511976667. [pp. 7, 9]

[22] Zdenek Hradil. Quantum-state estimation. *Physical Review A*, 55(3):R1561, 1997. doi:10.1103/PhysRevA.55.R1561. [p. 7]

[23] Matteo Paris and Jaroslav Rehacek. *Quantum State Estimation*, volume 649. Springer Science & Business Media, 2004. doi:10.1007/b98673. [p. 7]

[24] Robin Blume-Kohout. Optimal, reliable estimation of quantum states. *New Journal of Physics*, 12(4):043034, apr 2010. doi:10.1088/1367-2630/12/4/043034. [p. 7]

[25] K. Banaszek, M. Cramer, and D. Gross. Focus on quantum tomography. *New Journal of Physics*, 15(12):125020, dec 2013. doi:10.1088/1367-2630/15/12/125020. [p. 7]

[26] David Gross, Yi-Kai Liu, Steven T. Flammia, Stephen Becker, and Jens Eisert. Quantum State Tomography via Compressed Sensing. *Phys. Rev. Lett.*, 105:150401, Oct 2010. doi:10.1103/PhysRevLett.105.150401. [p. 7]

[27] Steven T. Flammia, David Gross, Yi-Kai Liu, and Jens Eisert. Quantum tomography via compressed sensing: error bounds, sample complexity and efficient estimators. *New Journal of Physics*, 14(9):095022, sep 2012. doi:10.1088/1367-2630/14/9/095022. [p. 7]

[28] Takanori Sugiyama, Peter S. Turner, and Mio Murao. Precision-Guaranteed Quantum Tomography. *Phys. Rev. Lett.*, 111:160406, Oct 2013. doi:10.1103/PhysRevLett.111.160406. [p. 7]

[29] Richard Kueng, Huangjun Zhu, and David Gross. Low rank matrix recovery from Clifford orbits. *arXiv preprint arXiv:1610.08070*, 2016. doi:10.48550/arXiv.1610.08070. [p. 7]

[30] Richard Kueng, Holger Rauhut, and Ulrich Terstiege. Low rank matrix recovery from rank one measurements. *Applied and Computational Harmonic Analysis*, 42(1):88–116, 2017. doi:10.1016/j.acha.2015.07.007. [p. 7]

[31] M Guţă, J. Kahn, R. Kueng, and J. A. Tropp. Fast state tomography with optimal error bounds. *Journal of Physics A: Mathematical and Theoretical*, 53(20):204001, apr 2020. doi:10.1088/1751-8121/ab8111. [p. 7]

[32] Marcus Cramer, Martin B. Plenio, Steven T. Flammia, Rolando Somma, David Gross, Stephen D. Bartlett, Olivier Landon-Cardinal, David Poulin, and Yi-Kai Liu. Efficient quantum state tomography. *Nature communications*, 1(1):1–7, 2010. doi:10.1038/ncomms1147. [p. 7]

[33] B.P. Lanyon, C. Maier, Milan Holzäpfel, Tillmann Baumgratz, C Hempel, P Jurcevic, Ish Dhand, A.S. Buyskikh, A.J. Daley, Marcus Cramer, et al. Efficient tomography of a quantum many-body system. *Nature Physics*, 13(12):1158–1162, 2017. doi:10.1038/nphys4244. [p. 7]

[34] Olivier Landon-Cardinal and David Poulin. Practical learning method for multi-scale entangled states. *New Journal of Physics*, 14(8):085004, aug 2012. doi:10.1088/1367-2630/14/8/085004. [p. 7]

[35] Juan Carrasquilla, Giacomo Torlai, Roger G. Melko, and Leandro Aolita. Reconstructing quantum states with generative models. *Nature Machine Intelligence*, 1(3):155–161, 2019. doi:10.1038/s42256-019-0028-1. [pp. 7, 8]

[36] Xun Gao and Lu-Ming Duan. Efficient representation of quantum many-body states with deep neural networks. *Nature communications*, 8(1):1–6, 2017. doi:10.1038/s41467-017-00705-2. [p. 7]

[37] Jordan Cotler and Frank Wilczek. Quantum overlapping tomography. *Phys. Rev. Lett.*, 124:100401, Mar 2020. doi:10.1103/PhysRevLett.124.100401. [p. 7]

[38] Scott Aaronson and Guy N. Rothblum. Gentle Measurement of Quantum States and Differential Privacy. In *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*, pages 322–333, 2019. doi:10.1145/3313276.3316378. [p. 8]

[39] Costin Bădescu and Ryan O'Donnell. Improved Quantum Data Analysis. In *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*, pages 1398–1411, 2021. doi:10.1145/3406325.3451109. [p. 8]

[40] Abhinav Kandala, Antonio Mezzacapo, Kristan Temme, Maika Takita, Markus Brink, Jerry M Chow, and Jay M. Gambetta. Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets. *Nature*, 549(7671):242–246, 2017. doi:10.1038/nature23879. [p. 8]

[41] Vladyslav Verteletskyi, Tzu-Ching Yen, and Artur F. Izmaylov. Measurement optimization in the variational quantum eigensolver using a minimum clique cover. *The Journal of Chemical Physics*, 152(12):124114, 2020. doi:10.1063/1.5141458. [p. 8]

[42] Artur F. Izmaylov, Tzu-Ching Yen, Robert A. Lang, and Vladyslav Verteletskyi. Unitary partitioning approach to the measurement problem in the variational quantum eigensolver method. *Journal of Chemical Theory and Computation*, 16(1):190–195, 2019. doi:10.1021/acs.jctc.9b00791. [p. 8]

[43] Andrew Zhao, Andrew Tranter, William M. Kirby, Shu Fay Ung, Akimasa Miyake, and Peter J. Love. Measurement reduction in variational quantum algorithms. *Physical Review A*, 101(6):062322, 2020. doi:10.1103/PhysRevA.101.062322. [p. 8]

[44] Guoming Wang, Dax Enshan Koh, Peter D. Johnson, and Yudong Cao. Minimizing Estimation Runtime on Noisy Quantum Computers. *PRX Quantum*, 2:010346, Mar 2021. doi:10.1103/PRXQuantum.2.010346. [p. 8]

[45] Dax Enshan Koh, Guoming Wang, Peter D. Johnson, and Yudong Cao. Foundations for Bayesian inference with engineered likelihood functions for robust amplitude estimation. *Journal of Mathematical Physics*, 63:052202, 2022. doi:10.1063/5.0042433. [p. 8]

[46] Jérôme F. Gonthier, Maxwell D. Radin, Corneliu Buda, Eric J. Doskocil, Clena M. Abuan, and Jhonathan Romero. Identifying challenges towards practical quantum advantage through resource estimation: the measurement roadblock in the variational quantum eigensolver. *arXiv preprint arXiv:2012.04001*, 2020. doi:10.48550/arXiv.2012.04001. [p. 8]

[47] Andrew Zhao, Nicholas C. Rubin, and Akimasa Miyake. Fermionic Partial Tomography via Classical Shadows. *Phys. Rev. Lett.*, 127:110504, Sep 2021. doi:10.1103/PhysRevLett.127.110504. [p. 8]

[48] Kianna Wan, William J. Huggins, Joonho Lee, and Ryan Babbush. Matchgate Shadows for Fermionic Quantum Simulation. *arXiv preprint arXiv:2207.13723*, 2022. doi:10.48550/arXiv.2207.13723. [p. 8]

[49] Bryan O'Gorman. Fermionic tomography and learning. *arXiv preprint arXiv:2207.14787*, 2022. doi:10.48550/arXiv.2207.14787. [p. 8]

[50] Charles Hadfield, Sergey Bravyi, Rudy Raymond, and Antonio Mezzacapo. Measurements of Quantum Hamiltonians with Locally-Biased Classical Shadows. *Communications in Mathematical Physics*, 391(3):951–967, 2022. doi:10.1007/s00220-022-04343-8. [pp. 8, 34]

[51] Andreas Elben, Richard Kueng, Hsin-Yuan Robert Huang, Rick van Bijnen, Christian Kokail, Marcello Dalmonte, Pasquale Calabrese, Barbara Kraus, John Preskill, Peter Zoller, et al. Mixed-State Entanglement from Local Randomized Measurements. *Physical Review Letters*, 125(20):200501, 2020. doi:10.1103/PhysRevLett.125.200501. [p. 8]

[52] G.I. Struchalin, Ya. A. Zagorovskii, E.V. Kovlakov, S.S. Straupe, and S.P. Kulik. Experimental Estimation of Quantum State Properties from Classical Shadows. *PRX Quantum*, 2:010307, Jan 2021. doi:10.1103/PRXQuantum.2.010307. [pp. 8, 34]

[53] Dax Enshan Koh and Sabee Grewal. Classical shadows with noise. *arXiv preprint arXiv:2011.11580v1*, 2020. [p. 8]

[54] Robin Harper, Steven T. Flammia, and Joel J. Wallman. Efficient learning of quantum noise. *Nature Physics*, 16(12):1184–1188, 2020. doi:10.1038/s41567-020-0992-8. [p. 8]

[55] Guangxi Li, Zhixin Song, and Xin Wang. VSQL: Variational shadow quantum learning for classification. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(9):8357–8365, May 2021. [p. 8]

[56] Joseph M. Lukens, Kody J. H. Law, and Ryan S. Bennink. A Bayesian analysis of classical shadows. *npj Quantum Inf.*, 7(113):1–10, Jul 2021. doi:10.1038/s41534-021-00447-6. [p. 8]

[57] Roy J. Garcia, You Zhou, and Arthur Jaffe. Quantum scrambling with classical shadows. *Phys. Rev. Research*, 3:033155, Aug 2021. doi:10.1103/PhysRevResearch.3.033155. [p. 8]

[58] Hong-Ye Hu and Yi-Zhuang You. Hamiltonian-driven shadow tomography of quantum states. *Phys. Rev. Research*, 4:013054, Jan 2022. doi:10.1103/PhysRevResearch.4.013054. [p. 8]

[59] Antoine Neven, Jose Carrasco, Vittorio Vitale, Christian Kokail, Andreas Elben, Marcello Dalmonte, Pasquale Calabrese, Peter Zoller, Benoît Vermersch, Richard Kueng, et al. Symmetry-resolved entanglement detection using partial transpose moments. *npj Quantum Inf.*, 7(152):1–12, Oct 2021. doi:10.1038/s41534-021-00487-y. [p. 8]

[60] Hsin-Yuan Huang, Richard Kueng, and John Preskill. Efficient estimation of Pauli observables by derandomization. *Phys. Rev. Lett.*, 127:030503, Jul 2021. doi:10.1103/PhysRevLett.127.030503. [p. 8]

[61] Atithi Acharya, Siddhartha Saha, and Anirvan M. Sengupta. Shadow tomography based on informationally complete positive operator-valued measure. *Phys. Rev. A*, 104:052418, Nov 2021. doi:10.1103/PhysRevA.104.052418. [p. 8]

[62] Stefan Hillmich, Charles Hadfield, Rudy Raymond, Antonio Mezzacapo, and Robert Wille. Decision Diagrams for Quantum Measurements with Shallow Circuits. In *2021*

*IEEE International Conference on Quantum Computing and Engineering (QCE)*, pages 24–34. IEEE, 2021. doi:10.1109/QCE52317.2021.00018. [p. 8]

[63] Charles Hadfield. Adaptive Pauli Shadows for Energy Estimation. *arXiv preprint arXiv:2105.12207*, 2021. doi:10.48550/arXiv.2105.12207. [p. 8]

[64] Bujiao Wu, Jinzhao Sun, Qi Huang, and Xiao Yuan. Overlapped grouping measurement: A unified framework for measuring quantum states. *arXiv preprint arXiv:2105.13091*, 2021. doi:10.48550/arXiv.2105.13091. [p. 8]

[65] Aniket Rath, Cyril Branciard, Anna Minguzzi, and Benoît Vermersch. Quantum Fisher information from randomized measurements. *Phys. Rev. Lett.*, 127:260501, Dec 2021. doi:10.1103/PhysRevLett.127.260501. [p. 8]

[66] Ting Zhang, Jinzhao Sun, Xiao-Xu Fang, Xiao-Ming Zhang, Xiao Yuan, and He Lu. Experimental quantum state measurement with classical shadows. *Phys. Rev. Lett.*, 127:200501, Nov 2021. doi:10.1103/PhysRevLett.127.200501. [pp. 8, 34]

[67] Hsin-Yuan Huang, Richard Kueng, Giacomo Torlai, Victor V. Albert, and John Preskill. Provably efficient machine learning for quantum many-body problems. *arXiv preprint arXiv:2106.12627*, 2021. doi:10.48550/arXiv.2106.12627. [p. 8]

[68] William J. Huggins, Bryan A. O'Gorman, Nicholas C. Rubin, David R. Reichman, Ryan Babbush, and Joonho Lee. Unbiasing fermionic quantum Monte Carlo with a quantum computer. *Nature*, 603(7901):416–420, Mar 2022. doi:10.1038/s41586-021-04351-z. [p. 8]

[69] Hong-Ye Hu, Soonwon Choi, and Yi-Zhuang You. Classical Shadow Tomography with Locally Scrambled Quantum Dynamics. *arXiv preprint arXiv:2107.04817*, 2021. doi:10.48550/arXiv.2107.04817. [p. 8]

[70] Steven T. Flammia. Averaged circuit eigenvalue sampling. *arXiv preprint arXiv:2108.05803*, 2021. doi:10.48550/arXiv.2108.05803. [p. 8]

[71] Ryan Levy, Di Luo, and Bryan K. Clark. Classical Shadows for Quantum Process Tomography on Near-term Quantum Computers. *arXiv preprint arXiv:2110.02965*, 2021. doi:10.48550/arXiv.2110.02965. [p. 8]

[72] Jonathan Kunjummen, Minh C. Tran, Daniel Carney, and Jacob M. Taylor. Shadow process tomography of quantum channels. *arXiv preprint arXiv:2110.03629*, 2021. doi:10.48550/arXiv.2110.03629. [p. 8]

[73] Jonas Helsen, Marios Ioannou, Ingo Roth, Jonas Kitzinger, Emilio Onorati, Albert H. Werner, and Jens Eisert. Estimating gate-set properties from random sequences. *arXiv preprint arXiv:2110.13178*, 2021. doi:10.48550/arXiv.2110.13178. [p. 8]

[74] Sitan Chen, Jordan Cotler, Hsin-Yuan Huang, and Jerry Li. Exponential Separations Between Learning With and Without Quantum Memory. In *2021 IEEE 62nd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 574–585, 2022. doi:10.1109/FOCS52979.2021.00063. [p. 8]

[75] Simone Notarnicola, Andreas Elben, Thierry Lahaye, Antoine Browaeys, Simone Montangero, and Benoit Vermersch. A randomized measurement toolbox for Rydberg quantum technologies. *arXiv preprint arXiv:2112.11046*, 2021. doi:10.48550/arXiv.2112.11046. [p. 8]

[76] Stefan H. Sack, Raimel A. Medina, Alexios A. Michailidis, Richard Kueng, and Maksym Serbyn. Avoiding barren plateaus using classical shadows. *PRX Quantum*, 3:020365, Jun 2022. doi:10.1103/PRXQuantum.3.020365. [p. 8]

[77] Kaifeng Bu, Dax Enshan Koh, Roy J. Garcia, and Arthur Jaffe. Classical shadows with Pauli-invariant unitary ensembles. *arXiv preprint arXiv:2202.03272*, 2022. doi:10.48550/arXiv.2202.03272. [p. 8]

[78] Max McGinley, Sebastian Leontica, Samuel J. Garratt, Jovan Jovanovic, and Steven H. Simon. Quantifying information scrambling via classical shadow tomography on programmable quantum simulators. *arXiv preprint arXiv:2202.05132*, 2022. doi:10.48550/arXiv.2202.05132. [p. 8]

[79] Lu Liu, Ting Zhang, Xiao Yuan, and He Lu. Experimental Investigation of Quantum Uncertainty Relations With Classical Shadows. *Frontiers in Physics*, 10, 2022. doi:10.3389/fphy.2022.873810. [pp. 8, 34]

[80] Joseph M. Lukens, Kody J. H. Law, and Ryan S. Bennink. Classical shadows and Bayesian mean estimation: a comparison. In *Conference on Lasers and Electro-Optics*, page FW3N.3. Optical Society of America, 2021. doi:10.1364/CLEO_QELS.2021.FW3N.3. [p. 8]

[81] Angus Lowe. Learning Quantum States Without Entangled Measurements. Master's thesis, University of Waterloo, 2021. [p. 8]

[82] Hsin-Yuan Huang. Learning quantum states from their classical shadows. *Nat. Rev. Phys.*, 4(2):81, Feb 2022. doi:10.1038/s42254-021-00411-5. [p. 8]

[83] Hong-Ye Hu, Ryan LaRose, Yi-Zhuang You, Eleanor Rieffel, and Zhihui Wang. Logical shadow tomography: Efficient estimation of error-mitigated observables. *arXiv preprint arXiv:2203.07263*, 2022. doi:10.48550/arXiv.2203.07263. [p. 8]

[84] Alireza Seif, Ze-Pei Cian, Sisi Zhou, Senrui Chen, and Liang Jiang. Shadow Distillation: Quantum Error Mitigation with Classical Shadows for Near-Term Quantum Processors. *arXiv preprint arXiv:2203.07309*, 2022. doi:10.48550/arXiv.2203.07309. [p. 8]

[85] Andreas Elben, Steven T. Flammia, Hsin-Yuan Huang, Richard Kueng, John Preskill, Benoît Vermersch, and Peter Zoller. The randomized measurement toolbox. *arXiv preprint arXiv:2203.11374*, 2022. doi:10.48550/arXiv.2203.11374. [p. 8]

[86] Gregory Boyd and Bálint Koczor. Training variational quantum circuits with CoVaR: covariance root finding with classical shadows. *arXiv preprint arXiv:2204.08494*, 2022. doi:10.48550/arXiv.2204.08494. [p. 8]

[87] H. Chau Nguyen, Jan Lennart Bönsel, Jonathan Steinberg, and Otfried Gühne. Optimising shadow tomography with generalised measurements. *arXiv preprint arXiv:2205.08990*, 2022. doi:10.48550/arXiv.2205.08990. [p. 8]

[88] Luuk Coopmans, Yuta Kikuchi, and Marcello Benedetti. Predicting Gibbs State Expectation Values with Pure Thermal Shadows. *arXiv preprint arXiv:2206.05302*, 2022. doi:10.48550/arXiv.2206.05302. [p. 8]

[89] Saumya Shivam, C. W. von Keyserlingk, and S. L. Sondhi. On Classical and Hybrid Shadows of Quantum States. *arXiv preprint arXiv:2206.06616*, 2022. doi:10.48550/arXiv.2206.06616. [p. 8]

[90] Daniel McNulty, Filip B. Maciejewski, and Michał Oszmaniec. Estimating Quantum Hamiltonians via Joint Measurements of Noisy Non-Commuting Observables. *arXiv preprint arXiv:2206.08912*, 2022. doi:10.48550/arXiv.2206.08912. [p. 8]

[91] Suguru Endo, Zhenyu Cai, Simon C. Benjamin, and Xiao Yuan. Hybrid quantum-classical algorithms and quantum error mitigation. *Journal of the Physical Society of Japan*, 90(3):032001, 2021. doi:10.7566/JPSJ.90.032001. [p. 8]

[92] Austin G. Fowler, Matteo Mariantoni, John M. Martinis, and Andrew N. Cleland. Surface codes: Towards practical large-scale quantum computation. *Physical Review A*, 86(3):032324, 2012. doi:10.1103/PhysRevA.86.032324. [p. 8]

[93] Earl T. Campbell, Barbara M. Terhal, and Christophe Vuillot. Roads towards fault-tolerant universal quantum computation. *Nature*, 549(7671):172–179, 2017. doi:10.1038/nature23460. [p. 8]

[94] Ying Li and Simon C. Benjamin. Efficient Variational Quantum Simulator Incorporating Active Error Minimization. *Phys. Rev. X*, 7:021050, Jun 2017. doi:10.1103/PhysRevX.7.021050. [p. 8]

[95] Kristan Temme, Sergey Bravyi, and Jay M. Gambetta. Error Mitigation for Short-Depth Quantum Circuits. *Phys. Rev. Lett.*, 119:180509, Nov 2017. doi:10.1103/PhysRevLett.119.180509. [pp. 8, 9]

[96] Tudor Giurgica-Tiron, Yousef Hindy, Ryan LaRose, Andrea Mari, and William J. Zeng. Digital zero noise extrapolation for quantum error mitigation. In *2020 IEEE International Conference on Quantum Computing and Engineering (QCE)*, pages 306–316, 2020. doi:10.1109/QCE49297.2020.00045. [p. 8]

[97] Piotr Czarnik, Andrew Arrasmith, Patrick J. Coles, and Lukasz Cincio. Error mitigation with Clifford quantum-circuit data. *Quantum*, 5:592, November 2021. doi:10.22331/q-2021-11-26-592. [p. 8]

[98] Jarrod R. McClean, Mollie E. Kimchi-Schwartz, Jonathan Carter, and Wibe A. de Jong. Hybrid quantum-classical hierarchy for mitigation of decoherence and determination of excited states. *Phys. Rev. A*, 95:042308, Apr 2017. doi:10.1103/PhysRevA.95.042308. [p. 8]

[99] Suguru Endo, Simon C. Benjamin, and Ying Li. Practical quantum error mitigation for near-future applications. *Phys. Rev. X*, 8:031027, Jul 2018. doi:10.1103/PhysRevX.8.031027. [pp. 8, 9]

[100] John Watrous. *The Theory of Quantum Information*. Cambridge University Press, 2018. doi:10.1017/9781316848142. [pp. 9, 10, 12, 18]

[101] Sepehr Nezami and Michael Walter. Multipartite entanglement in stabilizer tensor networks. *Phys. Rev. Lett.*, 125:241602, Dec 2020. doi:10.1103/PhysRevLett.125.241602. [p. 11]

[102] Fernando G. S. L. Brandao and Michal Horodecki. Exponential Quantum Speed-ups are Generic. *Quantum Inf. Comput.*, 13(11&12):901–924, 2013. doi:10.26421/QIC13.11-12-1. [p. 11]

[103] Adam Bouland, Joseph F. Fitzsimons, and Dax Enshan Koh. Complexity Classification of Conjugated Clifford Circuits. In Rocco A. Servedio, editor, *33rd Computational Complexity Conference (CCC 2018)*, volume 102 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 21:1–21:25, Dagstuhl, Germany, 2018. Schloss Dagstuhl–Leibniz-Zentrum für Informatik. doi:10.4230/LIPIcs.CCC.2018.21. [p. 11]

[104] Rawad Mezher, Joe Ghalbouni, Joseph Dgheim, and Damian Markham. Efficient approximate unitary t-designs from partially invertible universal sets and their application to quantum speedup. *arXiv preprint arXiv:1905.01504*, 2019. doi:10.48550/arXiv.1905.01504. [p. 11]

[105] Oleg Szehr, Frédéric Dupuis, Marco Tomamichel, and Renato Renner. Decoupling with unitary approximate two-designs. *New Journal of Physics*, 15(5):053022, 2013. doi:10.1088/1367-2630/15/5/053022. [p. 11]

[106] Andris Ambainis, Jan Bouda, and Andreas Winter. Nonmalleable encryption of quantum information. *Journal of Mathematical Physics*, 50(4):042106, 2009. doi:10.1063/1.3094756. [p. 11]

[107] Huangjun Zhu. Multiqubit Clifford groups are unitary 3-designs. *Physical Review A*, 96(6):062336, 2017. doi:10.1103/PhysRevA.96.062336. [p. 12]

[108] Joel J. Wallman. Randomized benchmarking with gate-dependent noise. *Quantum*, 2:47, January 2018. doi:10.22331/q-2018-01-29-47. [p. 23]

[109] Kevin Young, Stephen Bartlett, Robin J. Blume-Kohout, John King Gamble, Daniel Lobser, Peter Maunz, Erik Nielsen, Timothy James Proctor, Melissa Revelle, and Kenneth Michael Rudinger. Diagnosing and destroying non-Markovian noise. Technical report, Sandia National Lab. (SNL-CA), Livermore, CA (United States), 2020. doi:10.2172/1671379. [p. 34]

[110] Tilo Eggeling and Reinhard F. Werner. Separability properties of tripartite states with $U \otimes U \otimes U$ symmetry. *Physical Review A*, 63(4):042111, 2001. doi:10.1103/PhysRevA.63.042111. [p. 36]

[111] Peter D. Johnson and Lorenza Viola. Compatible quantum correlations: Extension problems for Werner and isotropic states. *Physical Review A*, 88(3):032323, 2013. doi:10.1103/PhysRevA.88.032323. [p. 36]