UC Berkeley
UC Berkeley Previously Published Works

Title

Shadow process tomography of quantum channels

Permalink

https://escholarship.org/uc/item/2j91p9k2

Journal

Physical Review A, 107(4)

ISSN

2469-9926

Authors

Kunjummen, Jonathan

Tran, Minh C

Carney, Daniel

et al.

Publication Date

2023-04-01

DOI

10.1103/physreva.107.042403

Shadow process tomography of quantum channels

Jonathan Kunjummen,[1, 2, *] Minh C. Tran,[3, 4, 1, 2] Daniel Carney,[5] and Jacob M. Taylor[1, 2]

[1]*Joint Center for Quantum Information and Computer Science,*
NIST/University of Maryland, College Park, Maryland 20742, USA
[2]*Joint Quantum Institute, NIST/University of Maryland, College Park, Maryland 20742, USA*
[3]*Center for Theoretical Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA*
[4]*Department of Physics, Harvard University, Cambridge, Massachusetts 02138, USA*
[5]*Physics Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA*
(Dated: May 27, 2023)

Quantum process tomography is a critical capability for building quantum computers, enabling quantum networks, and understanding quantum sensors. Like quantum state tomography, the process tomography of an arbitrary quantum channel requires a number of measurements that scales exponentially in the number of quantum bits affected. However, the recent field of shadow tomography, applied to quantum states, has demonstrated the ability to extract key information about a state with only polynomially many measurements. In this work, we apply the concepts of shadow state tomography to the challenge of characterizing quantum processes. We make use of the Choi isomorphism to directly apply rigorous bounds from shadow state tomography to shadow process tomography, and we find additional bounds on the number of measurements that are unique to process tomography. Our results, which include algorithms for implementing shadow process tomography, enable new techniques including evaluation of channel concatenation and the application of channels to shadows of quantum states. This provides a dramatic improvement for understanding large-scale quantum systems.

I. INTRODUCTION

Characterizing dynamical processes in quantum mechanics is a ubiquitous task. In the most general setting, a quantum process is described by a quantum channel–a linear map which takes a quantum state and maps it to another. In quantum computing, the execution of a quantum circuit is an attempt to faithfully produce a good approximation to a desired quantum channel, which represents the integrated effect of the specific experiment on the underlying quantum state. In quantum networking, quantum processes act during information transmission [1], characterizing uniquely quantum phenomena like the generation of entanglement while also leading to fundamental limitations to noiseless communication [2]. Beyond technological implications, understanding the properties of quantum channels underlies fundamental questions in disparate fields, including the characterization of entanglement properties of gravity [3–7].

Quantum channels can, in principle, dramatically change arbitrary quantum states, having an input and an output of a density matrix. Thus they require $O(d^4)$ different parameters to describe their action on a d-dimensional Hilbert space. Fully characterizing an arbitrary channel requires a number of measurements that is polynomial in the Hilbert space dimension, e.g., exponential in the number of qubits, even when channels are restricted to a smaller set such as unitary or nearly unitary channels [8]. Quantum process tomography is the detailed, and practical, exploration of how to do this as

effectively as possible [9–12]. Further difficulties in channel reconstruction also arise from the requirements that their action preserve the trace and positivity of density matrices [13].

Of course, the challenges that occur for quantum channels also occur for quantum states. In the past several years, there have been advances in reducing the number and type of measurements to better estimate properties of interest of quantum states by leveraging a concept from classical tomography: creating projective inverses of quasi-probability distributions. This extension to quantum states, denoted shadow quantum state tomography [14, 15], has already shown a dramatic reduction in the number of measurements necessary to efficiently estimate key properties, such as purity and expectation values of local operators [16–18]. Furthermore, these techniques use simple, easy to parallelize means of data storage and processing, and maximally leverage all experimental information per copy of the state estimated [19].

Here we apply the concept of shadow quantum state tomography (which we call shadow state tomography, dropping the 'quantum') to quantum channels. We show that many of the key results for shadow state tomography can be directly imported to quantum process tomography using 'process shadows'. Our approach leverages the Choi isomorphism [20–22], which has the natural interpretation of representing the action of a channel on one half of a maximally entangled state. However, key differences between inputs and outputs, as well as restrictions on Choi states, lead to a need to develop new measurement bounds, which we provide with a key theorem. Furthermore, we also show that shadows of quantum channels have additional useful properties, including their application to shadows of states, and the ability to concatenate

* jkunjumm@umd.edu

multiple shadow channels to create a new, quasi-shadow channel describing the more complex outcome.

Our paper begins with a brief introduction of state tomography and shadow state tomography, before leveraging the Choi isomorphism to provide an algorithm for shadow process tomography. We then prove the key theorem of this work describing rigorous bounds on measurement needs for accurately estimating the properties of a channel. Focusing on the case of preparing individual qubit Pauli eigenstates and measuring individual qubit outcomes, we show how to practically implement the shadow process tomography algorithm and consider the effects of applying a shadow process outcome to a shadow state. We find that both this case, and the case of concatenating channels together, leads to quasi-shadows with negative probabilities, and conjecture that an opportunity exists to reduce these quasi-shadows to maintain efficient performance. We conclude with numerical examples of our protocols applied to several different types of few qubit channels.

II. TRADITIONAL AND SHADOW TOMOGRAPHY

We start with a state ρ taken to be a density matrix for n qubits. Choosing a basis for operators $\{O_i\}$, we can write $\rho = \sum_i q_i O_i + \mathbb{I}/2^n$. Given this representation, we consider a complete set of traceless measurements $\{M_i\}$. Informational completeness tells us that we can find the q_i's given expectation values of the M_i's under ρ:

$$\langle M_i \rangle_\rho = \sum_j q_j \text{Tr}[M_i O_j] \tag{1}$$

If the matrix $\mathcal{M}_{ij} = \text{Tr}[M_i O_j]$ is invertible, then the linear system of equations defined above can be solved. In what follows, we will take the O's to be equal to the M's, and let them be relatively simple to implement, e.g., all Pauli strings over n qubits for a $d = 2^n$ dimensional space (the 'Clifford set'), or all single-qubit Paulis operators (the 'Pauli set'). In practice, the finite errors in expectation values due to sampling only a finite number of times leads to errors in the associated estimate of the density matrix, and nonlinear reconstruction techniques need to be employed to create physical (that is, completely positive and unit trace) density matrices that are also consistent with the measurement outcomes.

Here, we are instead focused on the question of how we can best use the measurement results, and use the language of shadow tomography [14, 15] to explore this question. In shadow tomography, we consider a scenario where finding $\mathcal{M}^{-1}$ is straightforward, and furthermore, where the sum can be done implicitly. Take a basis for Hilbert space $\{|b\rangle\}$ where b are n bit strings representing eigenstates of individual qubits in the Z basis. Let $\{U\}$ be a set of unitary operators and associating with each U there is a probability $p_U \geq 0$. We denote this ensemble

of unitaries by $\mathcal{U}$. We can define

$$\mathcal{M}_{\mathcal{U}}(\rho) = \mathbb{E}_U \sum_b \text{Tr}[U^\dagger |b\rangle\langle b| U \rho] \; U^\dagger |b\rangle\langle b| U. \tag{2}$$

We shall drop the subscript $\mathcal{U}$ if there is no ambiguity. We see that $\mathcal{M}$ is a linear operator. When $\mathcal{M}^{-1}$ exists, then the above equation has a simple interpretation. Let $p_{U,b} = p_U \text{Tr}[U^\dagger |b\rangle\langle b| U \rho]$. Then

$$\rho = \sum_{U,b} p_{U,b} \mathcal{M}^{-1}[U^\dagger |b\rangle\langle b| U] \tag{3}$$

up to normalization. That is, we take advantage of the linearity to do the inverse inside the sum. We see that the $p_{U,b}$ are positive numbers and can be interpreted as a probability. In general, we only sample over the sum on U and b, and thus the finite sample version of ρ only is expected to reproduce ρ in expectation.

For the case of single qubits and Pauli measurements, this inverse is straightforward to write down:

$$\mathcal{M}_{\mathcal{U}_\text{P}}^{-1}(A) = 3A - \text{Tr}(A)\mathbb{I}. \tag{4}$$

for any operator A. Here, we use $\mathcal{U}_\text{P}$ to denote the ensemble of uniformly random single-qubit Pauli operators. The inverse map for n-qubit Pauli measurements is simply $\mathcal{M}_{\mathcal{U}_\text{P}^{\otimes n}}^{-1} = (\mathcal{M}_{\mathcal{U}_\text{P}}^{-1})^{\otimes n}$. Similarly, for U taken from an ensemble $\mathcal{U}_\text{C}$ of arbitrary n-qubit Clifford circuits,

$$\mathcal{M}_{\mathcal{U}_\text{C}}^{-1}(A) = (2^n + 1)A - \text{Tr}(A)\mathbb{I}. \tag{5}$$

In practice, we take m copies of ρ, and for each copy, choose a random U (corresponding to setting a measurement basis) from our set. We then measure in the logical basis to get an outcome b. Let Q be the set of m such pairs of U and b and $\hat{\sigma}_j \equiv \mathcal{M}^{-1}(U_j^\dagger |b_j\rangle\langle b_j| U_j)$ for $j = 1, \ldots, m$. Then

$$\tilde{\sigma}_m = \frac{1}{m} \sum_{j=1}^m \hat{\sigma}_j \tag{6}$$

is a finite-sample estimator of ρ. In the limit of large sample size, $\lim_{m\to\infty} \tilde{\sigma}_m = \rho$.

Critically, work on this topic [15] has shown that only polynomially (in number of qubits) many such measurements are required to reproduce expectation values of certain observables. Given an observable O, let $\hat{o}_j = \text{Tr}(\hat{\sigma}_j O)$ for $i = 1, \ldots, m$. We then compute the median of means $\tilde{o}(N, K)$ by dividing Q into K subsets, each of size $N = m/K$, and calculating the median of the means of $\hat{o}_j$ in the subsets. The results of these are encompassed in the theorem from the above work, which we restate here:

Theorem 1 (State shadow tomography [15]). Given a set of density matrices Ω, a density matrix $\rho \in \Omega$, a unitary ensemble $\mathcal{U}$, a collection of M observables

$O_1, \ldots, O_M$, and any $\epsilon, \delta \in [0, 1]$,

$$K = 2\log(2M/\delta), \tag{7}$$

$$N = \frac{34}{\epsilon^2} \max_{1 \le \mu \le M} \min \left\{ \left\| O_\mu - \frac{\mathrm{Tr}(O_\mu)}{2^n} \mathbb{I} \right\|_{\mathrm{shadow},\mathcal{U},\Omega}^2, \right.$$
$$\left. \|O_\mu\|_{\mathrm{shadow},\mathcal{U},\Omega}^2 \right\}, \tag{8}$$

a set of $m = NK$ independent shadows are sufficient for

$$|\tilde{o}_\mu(N, K) - \mathrm{Tr}(O_\mu \rho)| \le \epsilon \quad \forall 1 \le \mu \le M \tag{9}$$

with probability at least $1 - \delta$. Here, the shadow norm of an observable O with respect to an unitary ensemble $\mathcal{U}$ and a set of density matrices is defined as

$$\|O\|_{\mathrm{shadow},\mathcal{U},\Omega}^2 := \max_{\sigma \in \Omega} \sum_{b \in \{0,1\}^n} \langle b| U\sigma U^\dagger |b\rangle$$
$$\times \langle b| U\mathcal{M}_\mathcal{U}^{-1}(O)U^\dagger |b\rangle^2, \tag{10}$$

where the maximization is over all density matrices σ in Ω. We will drop the subscript Ω if Ω is the set of all density matrices of n qubits.

In addition, Ref. [15] provides useful bounds on the shadow norms when $\mathcal{U}$ is either $\mathcal{U}_C$ or $\mathcal{U}_P^{\otimes n}$:

$$\|O\|_{\mathrm{shadow},\mathcal{U}_P^{\otimes n}}^2 \le 4^{\mathrm{supp}(O)} \|O\|^2, \tag{11}$$

$$\left\| O - \frac{\mathrm{Tr}(O)}{2^n}\mathbb{I} \right\|_{\mathrm{shadow},\mathcal{U}_C}^2 \le 3\mathrm{Tr}(O^2). \tag{12}$$

Equation (11) proves that the shadow tomography is efficient in estimating expectation values of local (few-qubit) observables when $\mathcal{U}$ is an ensemble of random local Pauli operators. On the other hand, if we choose U to be arbitrary Clifford circuits, Eq. (12) shows that the sample complexity scales with the rank of the observable and will be small for calculating, for example, the fidelity to a pure reference state. Therefore, depending on our purpose, we would choose a different unitary ensemble.

III. PROCESS SHADOWS

Just as state tomography builds an estimate of a density matrix, process tomography builds an estimate of a quantum channel. Recall that a channel $\mathcal{E}(\rho)$ can always be written using a Kraus representation as

$$\mathcal{E}(\rho) = \sum_{ij} \chi_{ij} K_i \rho K_j^\dagger \tag{13}$$

where $\{K_i\}$ is a complete and orthogonal (under the trace inner product) set of operators for the vector space of matrices of the same dimension $d \times d$ as ρ, and $\chi^\dagger = \chi$. The $\{K_i\}$ are often taken to be Pauli strings, though in principle other choices can be made. There is an additional constraint $\sum_{ij} \chi_{ij} K_j^\dagger K_i = \mathbb{I}$.

Given this representation, in process tomography we can consider a series of experiments in which we prepare the density matrix in some (known) state drawn from $\{\rho_i\}$ and find the expectation value of some set of observables $\{M_i\}$. Formally,

$$m_{ij} = \mathrm{Tr}[M_i \mathcal{E}(\rho_j)] = \sum_{kl} \chi_{kl} \mathrm{Tr}[M_i K_k \rho_j K_l^\dagger] \tag{14}$$

If we think of the vector space of matrices by combining i, j in a label a and k, l into a label b we see that if the matrix $\mathcal{K}_{ab} = \mathrm{Tr}[M_i K_k \rho_j K_l^\dagger]$ is invertible then we can recover χ_{kl}, just as we did for the density matrix in the previous section.

A. Choi isomorphism

We now ask: can the principles of shadow tomography be applied to process tomography? In order to apply the shadow tomography framework to a process $\mathcal{E}$, we make our lives conceptually easier through the use of (the Choi version of) the Choi-Jamiołkowski isomorphism to map a process into a density matrix. Specifically, we specialize to the case of finite dimension, where $\mathcal{E} : \mathcal{C}^{d \times d} \to \mathcal{C}^{d \times d}$ and is completely positive and trace-preserving. We define the unnormalized, maximally entangled state on a tensor product of two Hilbert spaces A and B each of dimension d

$$|\omega\rangle = \sum_{n=0}^{d-1} |n\rangle_A \otimes |n\rangle_B \tag{15}$$

(We remark that this state can be prepared by starting in the logical $|0\rangle_A \otimes |0\rangle_B$ state, applying Hadamard to each qubit in the first register, then applying pair-wise CNOTs from the ith qubit of the first register to the ith qubit of the second register.) We then define the Choi state

$$\eta = (\mathcal{I}_A \otimes \mathcal{E}_B)[|\omega\rangle\langle\omega|]. \tag{16}$$

We note that η is a density matrix (completely positive and of finite trace) up to a normalization factor d.

Given a Choi state η of dimension $d^2 \times d^2$ and an input density matrix ρ of size $d \times d$, we can find the result of $\mathcal{E}(\rho)$ as

$$\mathcal{E}(\rho) = \mathrm{Tr}_A[(\rho^T \otimes \mathbb{I}_B)\eta] \tag{17}$$

where we have taken the transpose of ρ with respect to the logical basis (that is, the basis in which we defined $|\omega\rangle$). This has the (simple) interpretation that given a Choi state, we can teleport ρ through $\mathcal{E}$ to find its result, as shown in Fig. 1.

To understand the teleportation interpretation, consider an arbitrary state $|\psi\rangle_C = \sum_n c_n |n\rangle$. Starting with
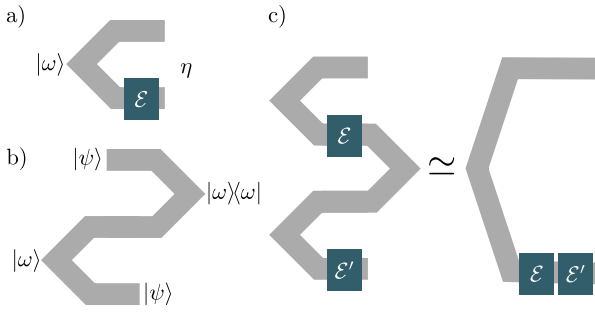
FIG. 1. (a) The Choi isomorphism in circuit form, (b) its use for state teleportation, and (c) its application for channel concatenation. An opening pair of thick grey lines represents creation of the maximally-entangled state $|\omega\rangle$. A closing pair indicates (successful) projection of the state onto $|\omega\rangle$.
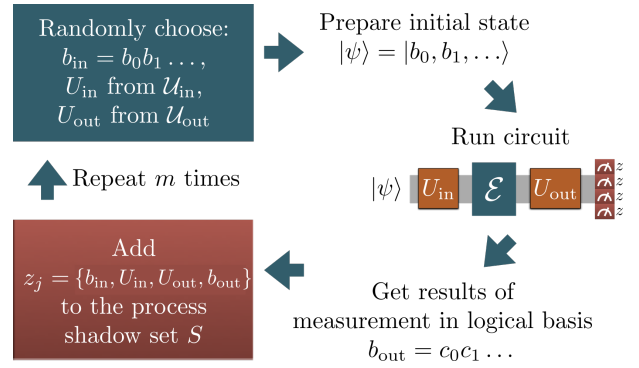

FIG. 2. Our shadow process tomography algorithm. We prepare a random product state in the logical basis, apply a random unitary U_{in} drawn from an ensemble $\mathcal{U}_{\text{in}}$, apply the channel $\mathcal{E}$, apply another random unitary U_{out} drawn from $\mathcal{U}_{\text{out}}$, and measure in the logical basis to obtain a bit string b_{out}. The combination $z = \{b_{\text{in}}, U_{\text{in}}, U_{\text{out}}, b_{\text{out}}\}$ forms a single-shot classical shadow of the channel.

$|\omega\rangle_{AB} |\psi\rangle_C$ and projecting onto $|\omega\rangle_{BC}$ yields:

$$
\langle\omega|_{BC} |\omega\rangle_{AB} |\psi\rangle_C
$$
$$
= \sum_{m,m',n} c_n \langle m|_B \langle m|_C |m'\rangle_A |m'\rangle_B |n\rangle_C \quad (18)
$$
$$
= \sum_{m,n} c_n \langle m|_C |m\rangle_A |n\rangle_C = \sum_m c_m |m\rangle_A \quad (19)
$$

Not surprisingly, the Choi state and its recovery of the channel is basis-specific, as the teleportation above makes explicit. However, its operational interpretation–that of state teleportation–we will find useful in applying shadow tomography concepts to process tomography.

B. Creating a process shadow

The Choi isomorphism provides a mapping from quantum channels to density matrices. Therefore in theory, we can perform process shadow tomography by applying shadow tomography on the Choi state. However, such a procedure is practically inefficient because we would need to prepare the highly entangled state $|\omega\rangle$ between *two* copies of the system. Instead, we use the teleportation interpretation of the Choi isomorphism and implement process shadow tomography using this procedure (see also Fig. 2):

1. Uniformly draw a random bit string $b_{\text{in}} \in \{0,1\}^n$ and prepare $|b_{\text{in}}\rangle$

2. Apply an unitary U_{in} drawn randomly from an ensemble $\mathcal{U}_{\text{in}}$. In particular, we will consider two possibilities for $\mathcal{U}_{\text{in}}$: random single-qubit Pauli rotations ($\mathcal{U}_{\text{P}}^{\otimes n}$) and random global Clifford ($\mathcal{U}_{\text{C}}$).

3. Apply the channel $\mathcal{E}$

4. Apply an unitary U_{out} drawn randomly from an ensemble $\mathcal{U}_{\text{out}}$, which is not necessary the same as

$\mathcal{U}_{\text{in}}$. Again, we will consider $\mathcal{U}_{\text{out}} = \mathcal{U}_{\text{P}}^{\otimes n}$ and $\mathcal{U}_{\text{out}} = \mathcal{U}_{\text{C}}$

5. Measure in the computational basis to obtain a bit string b_{out}.

6. Add the combination $z = \{b_{\text{in}}, U_{\text{in}}, U_{\text{out}}, b_{\text{out}}\}$ to the set ζ of shadow representations of the Choi state η (Eq. (16)).

7. Repeat the above m times.

Define

$$
|z\rangle = U_{\text{in}}^T |b_{\text{in}}\rangle \otimes U_{\text{out}}^\dagger |b_{\text{out}}\rangle . \quad (20)
$$

We show in Appendix A that the probability of getting a particular combination z can be written as

$$
P(z|\eta) = P(U_{\text{in}}, U_{\text{out}}, b_{\text{in}})\text{Tr}\big[|z\rangle\langle z|\eta\big] \quad (21)
$$

Then,

$$
\sum_z P(z|\eta)|z\rangle\langle z| = \mathcal{M}_{\mathcal{U}_{\text{in}}\otimes\mathcal{U}_{\text{out}}}(\eta) \quad (22)
$$

is a linear map of the Choi state. If $\mathcal{M}_{\mathcal{U}_{\text{in}}\otimes\mathcal{U}_{\text{out}}}$ is invertible, we define

$$
\hat{\zeta} = \mathcal{M}_{\mathcal{U}_{\text{in}}\otimes\mathcal{U}_{\text{out}}}^{-1}(|z\rangle\langle z|) \quad (23)
$$

as a single-shot shadow of η labeled by its parameters $z = \{b_{\text{in}}, \mathcal{U}_{\text{in}}, \mathcal{U}_{\text{out}}, b_{\text{out}}\}$. Clearly, taking the expectation over all possible such choices and outcomes for z will yield the original Choi state; that is, $\mathbb{E}_{U_{\text{in}}, U_{\text{out}}, b_{\text{in}}, b_{\text{out}}}\hat{\zeta} = \eta$.

Repeating the procedure above for m times, we obtain a collection of m shadows $\hat{\zeta}_1, \ldots, \hat{\zeta}_m$. The average of the collection, $\zeta = \frac{1}{m}\sum_{j=1}^m \hat{\zeta}_j$, provides an estimator for the Choi state $\eta = \lim_{m\to\infty}\zeta$.

In shadow tomography, an important metric is the sample complexity, i.e. the minimum m such that the estimator is close to correct values. Requiring the distance between the state estimator ζ and the Choi state η to be small is usually too restrictive (The sample complexity would scale polynomially with the Hilbert space dimension). Instead, we only ask that the estimator and the Choi state are close in computing the expectation values of certain observables.

Given a density matrix ρ and an observable O, we define $\hat{o} = \text{Tr}[\hat{\zeta}\rho^T \otimes O]$ as the single-shot estimator of $o = \text{Tr}[\mathcal{E}(\rho)O]$. Again, $\mathbb{E}\,\hat{o} = o$, where the mean is taken over all z combinations ($\mathcal{U}_{\text{in}}, \mathcal{U}_{\text{out}}, b_{\text{in}}$, and b_{out}). Considering a collection of m single-shot estimators $\hat{o}_1, \ldots, \hat{o}_m$, we define $\tilde{o}(\rho, N, K)$ as the median-of-means estimator, computed by dividing the collection into K subsets, each of size $N = m/K$, and taking the median of the means of the subsets.

Before we prove the sample complexity of process shadow tomography, we would like to note its qualitative differences from state shadow tomography. First, if we were to actually apply state shadow tomography on the Choi state, both b_{in} and b_{out} would have been measurement outcomes and their distributions depend on the Choi state η. In contrast, b_{in} in our procedure is drawn from an ensemble independent of η. Effectively, our procedure performs state shadow tomography on the Choi state and post-selects on the first copy of the system being in $|b_{\text{in}}\rangle$. We show in Appendix A that this distinction disappears when we draw b_{in} uniformly.

The second difference between state and process tomography is that the latter allows possibly different choices for $\mathcal{U}_{\text{in}}$ and $\mathcal{U}_{\text{out}}$. This enables four different mix-and-match scenarios when we restrict the ensembles to either $\mathcal{U}_{\text{C}}$ or $\mathcal{U}_{\text{P}}^{\otimes n}$. The following theorem, the central result of this work, estimates the sample complexity in the four scenarios, providing an analog of Theorem 1 for process shadow tomography.

Theorem 2. Let $O_1, \ldots, O_M$ be a collection of M operators and $\rho_1, \ldots, \rho_L$ be a collection of L other operators. Given $\epsilon, \delta \in [0, 1]$, let

$$K = 2\log(2ML/\delta), \tag{24}$$

$$N = \frac{34}{\epsilon^2} 4^n \max_{\substack{1 \le j \le M \\ 1 \le \ell \le M}} f_{\mathcal{U}_{\text{in}}}(\rho_\ell) f_{\mathcal{U}_{\text{out}}}(O_j), \tag{25}$$

where

$$f_{\mathcal{U}}(O) := \begin{cases} 4^{\text{supp}(O)} \|O\|^2 & \text{if } \mathcal{U} = \mathcal{U}_{\text{P}}^{\otimes n}, \\ \|\mathcal{S}(O)\| & \text{if } \mathcal{U} = \mathcal{U}_{\text{C}}, \end{cases} \tag{26}$$

$$\mathcal{S}(O) := 2\left\{[2\text{Tr}(O)^2 + \text{Tr}(O^2)]\mathbb{I} + 2\text{Tr}(O)O + 2O^2\right\}, \tag{27}$$

$\|.\|$ denotes the operator norm (also called the spectral norm) [15], and $\text{supp}(O)$ is the number of qubits O acts nontrivially on. Then, a collection of $m = NK$ shadows is sufficient to ensure that, with probability at least $1-\delta$,

the median of means $\hat{o}_j(\rho_k, N, K)$ is a good estimate of $\text{Tr}[\mathcal{E}(\rho_k)O_j]$, i.e.

$$|\text{Tr}[\mathcal{E}(\rho_k)O_j] - \hat{o}_j(\rho_k, N, K)| \le \epsilon, \tag{28}$$

for all $1 \le j \le M$ and $1 \le k \le L$.

Note that $\|\mathcal{S}(\rho)\| \le 14$ if ρ is a normalized density matrix and $\|\mathcal{S}(O)\| \le 14\,\text{rank}(O)^2\,\|O\|^2$ in general.

Proof. Using the Choi isomorphism, we can view the collections of M observables and L density matrices as ML "observables" of the form $2^n\rho_k^T \otimes O_j$ for $1 \le k \le L$ and $1 \le j \le M$. The leading factor of 2^n comes from the Choi state normalization, and will factor straightforwardly out of subsequent calculations to give the factor of 4^n in Eq. (25). We need to prove Theorem 2 for 4 different cases corresponding to $\mathcal{U}_{\text{in}}, \mathcal{U}_{\text{out}} \in \{\mathcal{U}_{\text{C}}, \mathcal{U}_{\text{P}}^{\otimes n}\}$.

First, we consider $\mathcal{U}_{\text{in}} = \mathcal{U}_{\text{out}} = \mathcal{U}_{\text{P}}^{\otimes n}$, which correspond to performing state shadow tomography on the Choi state using random Pauli measurements. This case has been analyzed in Ref. [15], which shows that

$$N = \frac{34}{\epsilon^2} \max_{\substack{1 \le \ell \le L \\ 1 \le j \le M}} 4^{\text{supp}(\rho_\ell \otimes O_j)} \left\|2^n\rho_\ell^T \otimes O_j\right\|^2 \tag{29}$$

$$= \frac{34}{\epsilon^2} 4^n \max_{\substack{1 \le \ell \le L \\ 1 \le j \le M}} 4^{\text{supp}(\rho_\ell \otimes O_j)} \|\rho_\ell\|^2 \|O_j\|^2 \tag{30}$$

is sufficient for the statement of Theorem 2.

To prove the other cases, we need the following lemma, which comes from the unitary 3-design property of $\mathcal{U}_{\text{C}}$ (see Appendix B for a proof):

Lemma 1. Given an operator O, we have

$$\sum_{b \in \{0,1\}^{\otimes n}} \mathbb{E}_{U \sim \mathcal{U}_{\text{C}}} U^\dagger |b\rangle\langle b| U \, \langle b|\, U\mathcal{M}_{\mathcal{U}_{\text{C}}}^{-1}(O) U^\dagger |b\rangle^2 \le 2\mathcal{S}(O). \tag{31}$$

With Lemma 1, we now prove Theorem 2 when $\mathcal{U}_{\text{in}} = \mathcal{U}_{\text{out}} = \mathcal{U}_{\text{C}}$. From Theorem 1, Theorem 2 holds if

$$N = \frac{34}{\epsilon^2} \left\|\rho^T \otimes O\right\|^2_{\text{shadow}, \mathcal{U}_{\text{C}} \otimes \mathcal{U}_{\text{C}}, \Omega_{\text{Choi}}} \tag{32}$$

$$= \frac{34}{\epsilon^2} \max_{\sigma \in \Omega_{\text{Choi}}} \sum_{b \in \{0,1\}^{2n}} \mathbb{E}_{U \sim \mathcal{U}_{\text{C}} \otimes \mathcal{U}_{\text{C}}} \text{Tr}\left[\sigma U^\dagger |b\rangle\langle b| U\right]$$
$$\times \langle b|\, U(\mathcal{M}_{\mathcal{U}_{\text{C}}}^{-1})^{\otimes 2}\left(2^n\rho^T \otimes O\right) U^\dagger |b\rangle^2 \tag{33}$$

$$\le \frac{34}{\epsilon^2} 4^n \max_{\sigma \in \Omega_{\text{Choi}}} \text{Tr}\left[\sigma\mathcal{S}(\rho^T) \otimes \mathcal{S}(O)\right] \tag{34}$$

$$\le \frac{34}{\epsilon^2} 4^n \left\|\mathcal{S}(\rho^T)\right\| \left\|\mathcal{S}(O)\right\|, \tag{35}$$

where Ω_{Choi} is the set of (normalized) Choi states and we have applied Lemma 1 twice. Note that $\text{Tr}(\sigma) = 1$. Therefore, $N = \frac{34}{\epsilon^2} 4^n \left\|\mathcal{S}(\rho^T)\right\| \left\|\mathcal{S}(O)\right\|$ is also sufficient for the statement of Theorem 2. Since $\left\|\mathcal{S}(\rho^T)\right\| = \left\|\mathcal{S}(\rho)\right\|$, Theorem 2 follows for $\mathcal{U}_{\text{in}} = \mathcal{U}_{\text{out}} = \mathcal{U}_{\text{C}}$.

Next, we consider the case $\mathcal{U}_{\mathrm{in}} = \mathcal{U}_{\mathrm{C}}$ and $\mathcal{U}_{\mathrm{out}} = \mathcal{U}_{\mathrm{P}}^{\otimes n}$. Again, by Theorem 1, Theorem 2 holds if

$$N = \frac{34}{\epsilon^2} \left\| 2^n \rho^T \otimes O \right\|_{\mathrm{shadow}, \mathcal{U}_{\mathrm{C}} \otimes \mathcal{U}_{\mathrm{P}}^{\otimes n}, \Omega_{\mathrm{Choi}}}^2 \tag{36}$$

$$\leq \frac{34}{\epsilon^2} 4^n \max_{\sigma \in \Omega_{\mathrm{Choi}}} \sum_{b_2} \mathbb{E}_{U_2 \sim \mathcal{U}_{\mathrm{P}}^{\otimes n}} \mathrm{Tr} \left[\sigma \left(\mathcal{S}(\rho^T) \otimes U_2^\dagger |b_2\rangle\langle b_2| U_2 \right) \right]$$
$$\times \langle b_2| U_2 \mathcal{M}_{\mathcal{U}_{\mathrm{P}}^{\otimes n}}^{-1}(O) U_2^\dagger |b_2\rangle^2 , \tag{37}$$

where we have also used Lemma 1 to evaluate the average over $U_1 \sim \mathcal{U}_{\mathrm{C}}$ and the sum over $b_1 \in \{0,1\}^n$. Let

$$\sigma_2 = \frac{1}{\|S(\rho^T)\|} \mathrm{Tr}_1 \left[\sigma \left(\mathcal{S}(\rho^T) \otimes \mathbb{I} \right) \right]. \tag{38}$$

It is straightforward to verify that $\sigma_2 \geq 0$ and $\mathrm{Tr}(\sigma_2) \leq 1$ for all $\sigma \in \Omega_{\mathrm{Choi}}$. Therefore, we can further upper bound the right-hand side of Eq. (37) by

$$\frac{34}{\epsilon^2} 4^n \left\| \mathcal{S}(\rho^T) \right\| \|O\|_{\mathrm{shadow}, \mathcal{U}_{\mathrm{P}}^{\otimes n}}$$
$$\leq \frac{34}{\epsilon^2} 4^n 4^{\mathrm{supp}(O)} \left\| \mathcal{S}(\rho^T) \right\| \|O\|, \tag{39}$$

leading to Theorem 2.

The case $\mathcal{U}_{\mathrm{in}} = \mathcal{U}_{\mathrm{P}}^{\otimes n}$ and $\mathcal{U}_{\mathrm{out}} = \mathcal{U}_{\mathrm{C}}$ follows the exact same steps with the roles of ρ^T and O exchanged. This completes the proof of Theorem 2. $\qquad\square$

We have now established concrete formal bounds for shadow process tomography which are similar in form to those for shadow state tomography. What does this mean for estimating quantities of interest?

One obvious difference between Theorem 2 and the analogous state shadows result is the daunting factor of 4^n in Eq. (25). Mathematically, this factor of 4^n follows from the fact that the Choi state of Eq. (16), which takes normalized input states to normalized output states via Eq. (17), has trace 2^n. In significant contrast to the state shadow case, this exponential factor affects even the use of semi-global Clifford measurements when ρ, O in Eq. (28) are rank-1 operators, i.e. state projectors. To explain the reason for the difference, in the state shadows case, if we prepare a (physical, unit trace) Choi state in the lab, the expectation value of the projector $\rho^T \otimes O$ can be estimated from randomized measurement with a cost independent of the system size. This total overlap includes a contribution from the exponentially small overlap of the input register state with ρ^T. By contrast, the expectation value in Eq. (28) corresponds to the probability of finding the channel-evolved state $\mathcal{E}(\rho)$ to be in state O; since a physical quantum channel will act without any initial probability of accepting or rejecting the input state, the probability of interest in Eq. (28) is exponentially larger than its physical state analogue.

Next, consider the choice for $\mathcal{U}_{\mathrm{in}}$. From Theorem 2, this choice will affect $f_{\mathcal{U}_{\mathrm{in}}}(\rho)$. If $\mathcal{U}_{\mathrm{in}} = \mathcal{U}_{\mathrm{C}}$ and ρ is a normalized density matrix,

$$f_{\mathcal{U}_{\mathrm{C}}}(\rho) = \|\mathcal{S}(\rho)\| \leq 14 \tag{40}$$

is bounded for all ρ. On the other hand, if $\mathcal{U}_{\mathrm{in}} = \mathcal{U}_{\mathrm{P}}^{\otimes n}$, $f_{\mathcal{U}_{\mathrm{P}}^{\otimes n}}(\rho)$ grows exponentially with the support size of ρ. So unless ρ is supported on only a few sites, we should always choose $\mathcal{U}_{\mathrm{in}} = \mathcal{U}_{\mathrm{C}}$ over $\mathcal{U}_{\mathrm{P}}^{\otimes n}$. Put in other terms, use of $\mathcal{U}_{\mathrm{P}}^{\otimes n}$ in the input space means that estimation of the effects of the channel work best on either high temperature states (ρ has a polynomial expansion in local operators in the inverse temperature) or on mostly depolarized inputs (such as most qubits set randomly and only a few polarized). On the other hand, the effects of the channel on highly entangled or highly polarized input states are best determined by using $\mathcal{U}_{\mathrm{C}}$ on the input.

Next, for the choice of $\mathcal{U}_{\mathrm{out}}$, we note that

$$f_{\mathcal{U}_{\mathrm{C}}}(O) \leq 14 \, \mathrm{rank}(O)^2 \, \|O\|^2 . \tag{41}$$

Similarly to the state shadow tomography, we would choose

- $\mathcal{U}_{\mathrm{out}} = \mathcal{U}_{\mathrm{C}}$ if O is a low-rank observable, e.g. the fidelity to a reference state, entanglement witness.

- $\mathcal{U}_{\mathrm{out}} = \mathcal{U}_{\mathrm{P}}^{\otimes n}$ is O is supported on a few qubits, e.g. local observables.

This becomes subtler and more important as we look to the next section, in which we consider a unique element of shadow process tomography: the ability to compose multiple channels together to create an estimate of the total channel thus developed.

Finally, there is a fundamental difference between the Choi state as concept and the Choi state as actual entangled state sent through a channel. Specifically, the latter case allows all options from shadow tomography. The former case only allows measurements that are not entangled between inputs and outputs. We will not remark on this additional power at this time, other than to say that in quantum sensing and related protocols the additional benefit of having access to the entangled bits can be substantial in, e.g., multiple interaction or applications of the channel for estimation [23].

IV. GETTING PRACTICAL: PAULI PREPARATION AND MEASUREMENTS

We now will consider some of the practical benefits and challenges of working with process shadows. We will focus on the case of the unitaries U being draw from single qubit operations $\mathcal{U}_{\mathrm{P}}^{\otimes n}$, as this is the most experimentally relevant regime at present. In contrast to the explicit teleportation-based tomographic concept implicit in the Choi state, we instead start by randomly sampling from both $\mathcal{U}_{\mathrm{P}}^{\otimes n}$ and b for the input strings. Let us denote the shadow process ζ as before. As may be obvious, adding to the list is straightforward by taking additional samples and then reweighting.

Given ζ and the corresponding set of z's S_m, we can estimate the action of the channel on qubit-qubit correlation functions, for example. Using S_m to create such estimates requires some care; borrowing from shadow state

tomography, we will in general suggest use of the median-of-means approach for any local estimators, as detailed in Ref. [15].

As this section specializes to the case of $\mathcal{U}_P^{\otimes n}$, it is convenient to define

$$\tau_{b,\mu} \equiv \mathcal{M}_{\mathcal{U}_P}^{-1}[|b_\mu\rangle\langle b_\mu|] = 3|b_\mu\rangle\langle b_\mu| - \mathbb{I} \qquad (42)$$

where $\mu = X, Y$, or Z and $b = \pm 1$.

A. Transition probabilities

One possible application of the above would be to estimation of transition probabilities. For transitions between computational-basis eigenstates, say from $|i\rangle \rightarrow |f\rangle$, this can be done by estimating the expectation value of the projector $O = |i\rangle\langle i| \otimes |f\rangle\langle f|$ in the Choi state η. A simple computation gives

$$\langle O\eta\rangle = \sum_a |\langle f|M_a|i\rangle|^2 = P(i \rightarrow f), \qquad (43)$$

where the M_a form a Kraus representation for the channel $\mathcal{E}(\rho) = \sum_a M_a \rho M_a^\dagger$. If the initial and final states $|i, f\rangle$ are localized to a small number k qubits and the rest of the system was considered in a high temperature (full mixed) state, then the projector O is a localized operator, and the random Pauli measurements should efficiently predict these transition probabilities. On the other hand, for an initial vacuum plus few excitation state, random Clifford measurements would yield reasonable bounds on estimation.

B. Multi-time correlation functions

Ref [15] notes that the Pauli measurement scheme is well suited for estimating spatial correlation functions, like the two point function $\langle \sigma_i^z \sigma_j^z \rangle = \text{Tr}[\rho \sigma_i^z \sigma_j^z]$ for sites i and j. Using shadows, we can extend this to calculation correlation functions in time between local operators: $\text{Tr}[\mathcal{E}(\rho A)B]$. Note, however, that the shadow norm squared for this correlation function goes as 4^k where k is the size of the support of $\rho A \otimes B$. As noted above, this adds a constraint that ρ be composed mostly of low-weight strings. As a result it is useful to turn emphasis away from the initial state and to consider the case $\rho = \mathbb{I}$ and unitary $\mathcal{E}$. In this case we don't consider the correlation function with respect to a particular state, rather we look learning about how operators evolve in the Heisenberg picture. For a unitary U corresponding to evolution for time t we can estimate functions like

$$\text{Tr}[\mathcal{E}(\rho A)B] \rightarrow \text{Tr}[UAU^\dagger B] = \text{Tr}[AB(t)] \qquad (44)$$

with a sample cost exponential in the support of $A \otimes B$. More details on this application (including to nonunitary channels) can be found in Appendix C.

C. Unitarity verification

Shadow tomography is capable of predicting functions which are both linear and non-linear in the quantum state. In particular, Ref. [15] gives an algorithm for estimating $\text{tr}(O\rho^2)$ for an arbitrary operator O. This proceeds by finding a linear operator $\tilde{O}$ such that $\text{Tr}(O\rho^2) = \text{Tr}(\tilde{O}\rho \otimes \rho)$, where the first trace is on a Hilbert space $\mathcal{H}$ while the second is on $\mathcal{H} \otimes \mathcal{H}$. For example, we can estimate the purity of a state ρ using $O = S$ the SWAP operator, because

$$\text{Tr}_\mathcal{H}\rho^2 = \sum_{ij} |\rho_{ij}|^2 = \text{Tr}_{\mathcal{H}\otimes\mathcal{H}}[S\rho \otimes \rho]. \qquad (45)$$

Applied to the Choi state of a given channel, this can be used to verify if the channel is unitary, i.e. has a single Kraus operator.

To see this, note that a channel $\mathcal{E}$ is unitary if and only if its Choi state,

$$\eta = \sum_{ij} |i\rangle\langle j| \otimes \mathcal{E}(|i\rangle\langle j|) \qquad (46)$$

is maximally entangled. (To see it in one direction: if $\mathcal{E}$ is unitary, one easily checks that $\eta^2 = \eta$, up to normalization. The converse is slightly more tedious but straightforward). From the definition, any Choi state has the property that the partial trace over the output system B gives the maximally mixed state for the input system A:

$$\eta_A := \text{Tr}_B\eta = \mathbb{I}_A. \qquad (47)$$

Therefore, if the total state η is pure up to normalization, $\text{Tr}\,\eta^2 = d^2$ (or equivalently $\eta^2 = d\eta$, where $d = \dim H_A$), then we can conclude that the Choi state is maximally entangled, and thus that the channel ϵ is unitary.

Unfortunately, the SWAP operator S is completely non-local: k qubits partitioned into two sets of $k/2$ qubits, the SWAP operator acts on all k qubits. Thus according to the estimates in [15], random Pauli measurements would require a number of samples of order 4^k to estimate $\text{Tr}\,\eta^2$. This is the same requirement as doing full process tomography. Random Clifford measurements likewise require a sample size scaling exponentially with the number of qubits. It would be extremely interesting to determine if there exists a measurement set that can produce non-exponential scaling, or to prove that no such set can exist.

This task has two somewhat disparate applications in quantum gravity. One is to the black hole information paradox. The complexity of decoding Hawking radiation to determine the initial, pre-black hole state is believed to be exponential in the logarithm of the Hilbert space dimension [24]. A more basic question is to simply check whether the formation and radiation process produces a unitary channel. In a very different regime, many recent proposals have been made to experimentally determine if Newtonian gravity can entangle meso-to-macroscopic

objects in a lab [3–7]. Understanding whether the channel generated by the gravitational interaction is unitary or not is crucially important in determining the implications of these experiments [25]. These experiments typically involve just a few low-dimensional systems (e.g., two qubits [4]), and so the above exponential scaling does not present a substantial difficulty.

D. Application to an incoming shadow state

What if we start with a shadow tomography estimator σ to the input state, ρ_{in}, which was derived from sampling over $\mathcal{U}_P^{\otimes n}$ using shadow state tomography? If we have the estimator for the Choi state ζ we need to calculate

$$\sigma_{\text{out}} \text{Tr}[\sigma_{in}^T \zeta] = \sum_{\mu,b,\mu',b',\nu,c} \prod_{ij} \text{Tr}[\tau_{b_j,\mu_j}^T \tau_{b'_j,\mu'_j}] \tau_{c_i,\nu_i}^{(i)} \quad (48)$$

where the sum is over the set S_m and the similar set R_k that defines σ. Note that the transpose operation on τ simply flips the sign of the Y-portion of the Pauli representation, analogous to complex conjugation, for this simple channel example.

Fortunately, summing up over sample pairs is relatively simple, and the final state involves a sum over only the output samples ν, c. This seems like we have a statement that shadows map to shadows. However, unlike our original shadows, this estimator for the output no longer has positive values in front of each τ product. Let us consider the quasiprobability term:

$$\frac{1}{2}\text{Tr}[\tau_{b,\mu}^T \tau_{b',\mu'}] = \begin{cases} 5/2, & \mu = \mu' \neq Y \text{ and } b = b' \\ -2, & \mu = \mu' \neq Y \text{ but } b \neq b' \\ -2, & \mu = \mu' = Y \text{ and } b = b' \\ 5/2, & \mu = \mu' = Y \text{ but } b \neq b' \\ 1/4, & \text{otherwise} \end{cases} \quad (49)$$

This points to an interesting future research direction: better understanding the emergence of an (effective) sign problem in the application of a shadow to a shadow; we discuss this in more detail at the end of this section.

E. Channel composition

Just as we can apply a shadow process to a shadow state, consider what happens if we have process shadows for two channels, $\mathcal{X}$ and $\mathcal{Y}$, each with Choi state shadows η_X, η_Y. Can we construct an approximation to the channel $\mathcal{Y} \circ \mathcal{X}$?

Using the teleportation-based interpretation of the Choi state, it suffices to project onto a maximally entangled state between the output register of X, B, and the input register of Y, A'. Using the Choi representation of both processes, with set S_m of b, μ, c, ν values for the

first channel and $S_{m'}$ of b', μ', c', ν' values for the second channel, we find immediately

$$\sigma_{\text{con}} = \sum_{(c,\nu,b,\mu),(b',\mu',c',\nu')} \prod_{ii'} \tau_{c,\nu}^{(i)} \tau_{c',\nu'}^{(i')} \text{Tr}[\tau_{b,\mu}^T \tau_{b',\mu'}] \quad (50)$$

F. Distribution for random states

Consider a random variable s drawn uniformly from $S = \{5/2, 1/4, 1/4, 1/4, 1/4, -2\}$. We are interested in properties of

$$\hat{\pi} = \Pi_i^N s_i. \quad (51)$$

This corresponds to composing two N-qubit channels with Pauli shadows according to Eqs. Eq. (49), Eq. (50) without any assuming any correlations between qubits. While the signed weights in Eq. (49) point to the existence of a sign problem, we want to investigate how important it is in practice to keep track of weights of small absolute value or of negative sign. First, we can ask what is the probability that k of the N samples are either $5/2$ or -2. We get

$$p_{\text{big}}(k) = \binom{N}{k}(1/3)^k (2/3)^{N-k}. \quad (52)$$

Then, given that we have k drawn uniformly from $5/2$ and -2, we want to know how many -2's we have, and if the overall number is even. The conditional averages are

$$p_{\text{odd}}(k) = \lim_{p \to 1/2} \sum_{j \in \text{odd}} \binom{k}{j}(p)^j (1/2)^{k-j} \quad (53)$$

$$= \lim_{p \to 1/2} \frac{1}{2}\left[(1/2 + p)^k - (1/2 - p)^k\right] = \frac{1}{2} \quad (54)$$

$$p_{\text{even}}(k) = \sum_{j \in \text{even}} \binom{k}{j}(1/2)^j (1/2)^{k-j} \quad (55)$$

$$= \lim_{p \to 1/2} \frac{1}{2}\left[(1/2 + p)^k + (1/2 - p)^k\right] = \frac{1}{2}. \quad (56)$$

From this we see that, for $k \geq 1$, we have an equal probability of having an even number of -2's and an odd number of -2's. We also have the average number of -2's is independent of being even or odd and is $k/2$.

The total probability over all S options for having a negative weight (an odd number of -2's) is

$$p_{\text{odd}} = \frac{1}{2}\left(1 - (2/3)^N\right). \quad (57)$$

What is the total probability of positive? We have the chance that no large values are chosen, $(2/3)^N$, and then we have the chance that some number are chosen, in which case we have a $1/2$ chance of being positive, giving

$$p_{\text{even}} = 1/2(1 + (2/3)^N), \quad (58)$$

as expected.

Thus we conclude that for random states, there is only a slight bias towards positive weights $\pi > 0$ as the number of qubits, N, becomes large. We can understand this bias as related to the probability of having no $|S| > 1$ values at all, e.g., $(2/3)^N$. We cannot hope to avoid the sign problem by having only a few negative weights.

However, while the weights are only slightly biased towards being positive, the amplitude of positive states is more heavily biased. We rewrite $\pi = (-1)^l A$ where $A > 0$, we can look at $a = \log A$ as a random variable, which is correlated with $l \in \{0, 1\}$. For $k > 0$ large values ($|S| > 1$) we have l evenly distributed and $\bar{a}(k)$ corresponds to $(N - k)\log 1/4 + \frac{k}{2}(\log 5/2 + \log 2)$. In the large N limit, a becomes normally distributed, which means that returning to the distribution of $|A|$, we have (except for a $(2/3)^N$ correction) a log normal distribution with log mean given by averaging $\bar{\bar{a}} = \sum_k \bar{a}(k)p_k \sim \frac{2N}{3}\log(1/4) + \frac{N}{6}\log 5$ and standard deviation $\sim \sqrt{N/6}\log 5$. The log normal distribution implies that a few, rare values will dominate the overall behavior of the weights. It is then worth investigating whether strategies based on importance sampling can accelerate the process of channel composition.

V. NUMERICS

We now explore the capabilities of classical shadow process tomography with numerical simulations, providing conceptual and practical tests of the theory provided before. There are a few new things to investigate compared to previous numerical experiments in state tomography. First, as discussed above, there is the effect of a particular input state playing the role of an observable and thus contributing to the shadow norm. Second, there is the effect of channel composition.

In order to generate different Choi states so we can test the properties of a wide variety of channels (unitary, nonunitary, etc.), we make use of the numerical quantum information packages `qi` and `qinf050` [26, 27]. The most significant use we make of these libraries is generating random unitary matrices which, using Stinespring dilation, is sufficient to generate arbitrary quantum channels. Below, in order to generate random nonunitary channels on n_{sys} qubits, we generate a random unitary U_{rand} on a system of $n_{sys} + n_{anc}$ qubits, where $n_{anc} = 2n_{sys}$ is enough to give a full rank channel. We define the jth Kraus operator K_j via

$$K_j = \text{Tr}_{anc}[U_{rand}|0\rangle\langle j| \otimes \mathbb{I}_{sys}] \qquad (59)$$

with $|j\rangle$ the jth ancilla basis state.

In this work, we work with systems of sufficiently small size that we can test convergence of the shadow representation of the Choi matrix to the full Choi matrix directly. As system size scales up, results from classical shadow state tomography suggest that shadow estimates
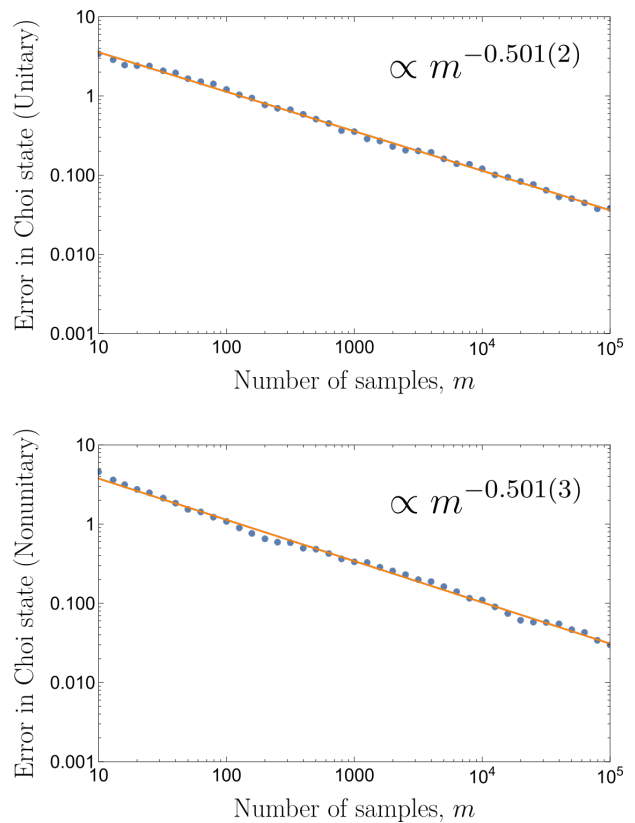

FIG. 3. Numerical demonstration of the effectiveness of the experimental protocol for Pauli string operations on input and output for two qubits. Shadow representation of the Choi state from generating a Pauli bit string, acting on it with the channel, and measuring in random Pauli basis does converge to Choi state. For our error metric we use the operator norm, the absolute value of the largest eigenvalue of $\eta - \zeta$. Convergence goes as one over square root of sample number for both unitary (top) and general (bottom) channels. Data (blue points) and trendline (orange) are plotted on a log-log scale and correspond to shadows of a fixed channel. Insets in top right show average convergence exponent for ten such random channels.

of local correlation functions will converge much more rapidly than the actual state estimate. We also therefore investigate the convergence properties of these correlation functions. Below, when we speak of convergence of our estimate of an operator rather than of an expectation value, we always mean with respect to the operator norm, i.e. our error ϵ in some estimate ζ of an operator η is $\epsilon = ||\zeta - \eta||$.

In our first set of numerical experiments we investigate the properties of a single channel. In Fig. 3, we calculate the error between our shadow representation of a Choi state corresponding to a random channel. We see the expected convergence behavior, with no significant difference between taking shadows of Choi states corresponding to unitary vs. nonunitary channels. In Fig. 4 we investigate the convergence properties of the output state resulting from tracing a specific input state against
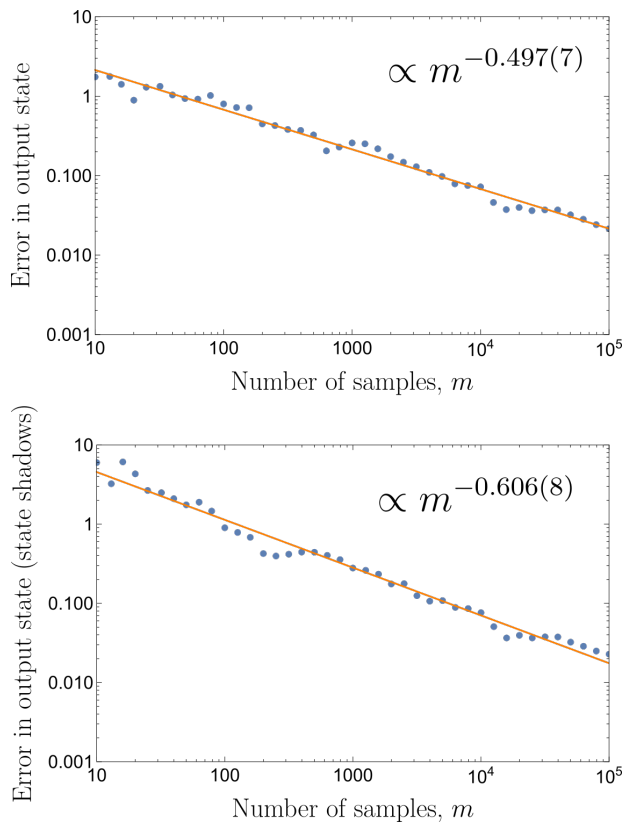
FIG. 4. Convergence properties of shadow estimate of channel acting on a specific input density matrix for two qubits. We show convergence of our estimate of the output density matrix to the true density matrix using both the full input state (top) and a shadow representation of it (bottom). Error in calculating the output state is again the operator norm. Data (blue points) and trendline (orange) are plotted on a log-log scale and correspond to shadows of a fixed channel. Insets in top right show average convergence exponent for ten such random channels.

the Choi matrix. We generate a random density matrix with respect to the Hilbert-Schmidt measure, and subject it to a random channel. We then compare the convergence of the shadow estimate of the output operator to the actual operator evolved under the channel, first using the full input density matrix and second using a shadow estimator for the input density matrix. In the latter case, we fix the number of shadow samples of the input density matrix to be the same as the number of channel samples.

In Fig. 5 we calculate the temporal correlation function

$$\text{Tr}[\mathcal{E}(\rho_{in}\sigma_1^x)\sigma_1^x] \qquad (60)$$

where $\rho_{in} = |+\rangle\langle+| \otimes \frac{\mathbb{I}}{2}$, as described in Appendix C. We use this input density matrix so we will have an operator made of low-weight Pauli strings, which is well suited to classical shadow estimation with local Pauli operations. We calculate the error by taking the absolute value of the difference between our correlation function calculated with shadows and the actual correlation function.
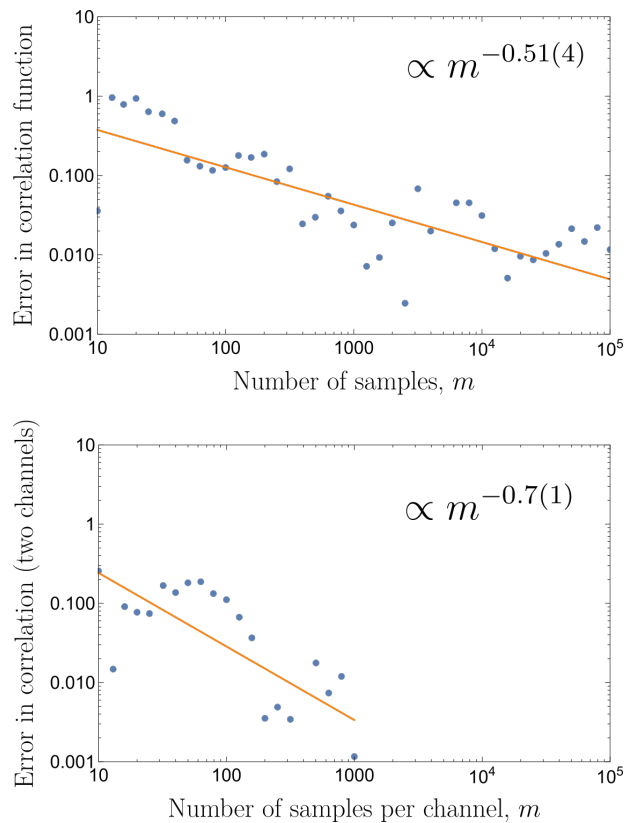
FIG. 5. Estimating multitime correlation functions for two qubits. Error in the correlation function is the absolute value of the difference between the shadow calculation and the true value. We do this for a single channel (top) and for the composition of two channels (bottom). Data (blue points) and trendline (orange) are plotted on a log-log scale and correspond to shadows of a fixed channel or pair of channels. Insets in top right show average convergence exponent for ten such random channels.

We then calculate the same correlation function, but this time for a channel which is the composition of two channels for which we have shadow representations. Mathematically, one can concatenate channels in the Choi representation by projecting the output of the first channel and the input of the second into the maximally entangled state, as shown in Fig. 1. We implement this procedure using the shadow representations of the two Choi states, using Eq. (50). We use the same number of samples m on each leg of the channel, so that when we concatenate, we do a sum over m^2 terms. In the above numerics, we identify convergence behavior by assuming the error ϵ decays with the number of samples m as a power law, i.e.

$$\epsilon \propto m^{-b} \qquad (61)$$

We estimate the convergence exponent b by doing linear regression on a log-log plot of the error versus number of samples, as the power law is determined by the slope of the line of fit. We find that errors in all cases with a single object represented by shadows scale as the number

of samples raised to the -1/2 power, as expected from our main theorem. Curiously, combining state and channel shadows or channel and channel shadows, on the other hand, appear to converge faster. However we note that in the case of correlation functions with composed channels, the convergence exponent varies significantly across realizations, and it is not clear that a power-law fit really captures the convergence behavior.

VI. OUTLOOK

We have proposed a method for characterizing quantum processes using shadow process tomography. We have proved rigorous bounds on the performance of this method and conducted numerical experiments to analyze the ability of shadow tomography to predict specific features of a quantum channel. We have discussed why this technique suffers from an exponential cost in system size that is not present in the state tomography case. Some theoretical questions remain. For instance, what is the rigorous performance guarantee on applying shadows of process to shadows of quantum states? If the channel we seek to characterize has special properties (e.g. unitarity), can we use compressed sensing techniques to use the measurement record more efficiently? Can we use importance sampling to compose shadow representations of channels more efficiently? There are also opportunities in thinking about how to apply shadow process tomography in the near term. Perhaps using shadow process tomography to carry out channel composition in a computer, it is possible to propagate quantum simulation processes out past the experimental limits on runtime. Finally, these techniques may be a useful tool in verifying fidelity of complicated, multi-qubit gates.

VII. ACKNOWLEDGMENTS

Note added: In preparation of this manuscript we became aware of related work by Levy et al. [28]. Our results and those of [28] were obtained independently. We also note related work in [29].

VIII. AUTHOR CONTRIBUTIONS

JK and MCT contributed equally to this work.

Appendix A: Effects of the probability distribution of b_{in}

In this section, we discuss the effect of the distribution of b_{in} on the process shadow tomography. As mentioned in the main text, if we were to perform the state shadow tomography on the Choi state η, the probability of obtaining a certain combination $z = \{b_{in}, b_{out}, U_{in}, U_{out}\}$ would depend on η as

$$P(z|\eta) = \frac{1}{2^n} P(U_{in}, U_{out}) \mathrm{Tr}(|z\rangle\langle z|\eta), \qquad \text{(A1)}$$

where $|z\rangle$ is the state defined in Eq. (20) and the factor $1/2^n$ accounts for the renormalization of η.

In contrast to shadow state tomography where the probability of b_{in} depends on η, we pick an η-independent probability distribution $P(b_{in})$ for b_{in} in our process shadow tomography. Therefore, the probability of obtaining a combination z is

$$P(z|\eta) = P(b_{in}, U_{in}, U_{out}) P(b_{out}|b_{in}, U_{in}, U_{out}, \eta)$$
$$= P(b_{in}, U_{in}, U_{out}) \frac{\mathrm{Tr}(|z\rangle\langle z|\eta)}{\mathrm{Tr}\left[\eta(U_{in}^T|b_{in}\rangle\langle b_{in}|U_{in}^*) \otimes \mathbb{I}\right]}, \quad \text{(A2)}$$

where the denominator on the right-hand side is a normalization factor for the state resulted from projecting $U_{in}^* \otimes U_{out}\, \eta\, (U_{in}^* \otimes U_{out})^\dagger$ onto the subspace where the first copy of the system is in the state $|b_{in}\rangle$.

Equations (A1) and (A2) are generally different. However, because η is a Choi state and the associated channel $\mathcal{E}$ is trace-preserving, the denominator on the right-hand side takes a trivial value:

$$\mathrm{Tr}\left[\eta(U_{in}^T|b_{in}\rangle\langle b_{in}|U_{in}^*) \otimes \mathbb{I}\right] = \mathrm{Tr}(\mathcal{E}(U_{in}|b_{in}\rangle\langle b_{in}|U_{in}^\dagger)) = 1$$
$$\text{(A3)}$$

for all b_{in}. With b_{in} being drawn uniformly at random ($P(b_{in}) = 1/2^n$), Eq. (A2) reduces to Eq. (A1) and we can treat b_{in} as if it were the result of performing state shadow tomography on the Choi state.

A different way to show this is to consider measurement of the input qubits according to the Pauli operators set by the string $P_A = \prod_i \mu^{(i)}$ which operates only the input bits. For Choi states, they will always have the expectation value

$$\langle P \rangle = \mathrm{Tr}[P_A \mathcal{I}_A \otimes \mathcal{E}_B |\omega\rangle\langle\omega|] \qquad\qquad \text{(A4)}$$
$$= \sum_{jmn} \mathrm{Tr}[P_A |m\rangle\langle n|] \mathrm{Tr}[K_j |m\rangle\langle n| K_j^\dagger] \qquad \text{(A5)}$$
$$= \sum_{mn} \mathrm{Tr}[P_A |m\rangle\langle n|] \mathrm{Tr}[\sum_j K_j^\dagger K_j |m\rangle\langle n|] \qquad \text{(A6)}$$
$$= \sum_m \mathrm{Tr}[P_A |m\rangle\langle m|] = \mathrm{Tr}[P_A] = 0, \qquad \text{(A7)}$$

where we write $\mathcal{E}_B(\rho) = \sum_j K_j \rho K_j^\dagger$ using the Kraus representation. We also use the identity $\sum_j K_j^\dagger K_j = \mathbb{I}$. Thus independent of the channel, there is no bias in the expectation value of any traceless observable on the input; by randomly choosing an input qubit state, we reproduce this unbiased input result.

Appendix B: Proof of Lemma 1

We provide a proof of Lemma 1 in this section.

Proof. Since $\mathcal{U}_C$ is unitary 3-design, we have the following property [30, Lemma 7]:

$$\mathbb{E}_{U \sim \mathcal{U}_C} U^\dagger |b\rangle\langle b| U = \frac{\mathbb{I}}{2^n}, \tag{B1}$$

$$\mathbb{E}_{U \sim \mathcal{U}_C} U^\dagger |b\rangle\langle b| U \, \langle b| \, UBU^\dagger \, |b\rangle^2$$
$$= \frac{\text{Tr}(B)^2 \mathbb{I} + 2\text{Tr}(B)B + 2B^2 + \text{Tr}(B^2)\mathbb{I}}{2^n(2^n+1)(2^n+2)}, \tag{B2}$$

for any operator B. Using Eq. (5) for $\mathcal{M}_{\mathcal{U}_C}^{-1}$ and the Cauchy-Schwarz inequality, we have

$$\sum_{b \in \{0,1\}^n} \mathbb{E}_{U \sim \mathcal{U}_C} U^\dagger |b\rangle\langle b| U \, \langle b| \, U \mathcal{M}_{\mathcal{U}_C}^{-1}(O) U^\dagger |b\rangle^2$$
$$\leq 2 \sum_{b \in \{0,1\}^{\otimes n}} \mathbb{E}_{U \sim \mathcal{U}_C} U^\dagger |b\rangle\langle b| U$$
$$\times \left[(2^n+1)^2 \, \langle b| \, UOU^\dagger \, |b\rangle^2 + \text{Tr}(O)^2 \right] \tag{B3}$$
$$\leq 2 \left[2\text{Tr}(O)^2 \mathbb{I} + 2\text{Tr}(O)O + 2O^2 + \text{Tr}(O^2)\mathbb{I} \right]. \tag{B4}$$

Therefore, Lemma 1 follows. $\qquad\square$

Appendix C: Calculating multitime correlation functions with classical shadows

We now examine how shadows can be used to estimate correlation functions. First, look at the correlator of the Pauli x operator on site 1, σ_1^x, and the time-delayed Pauli y operator on site 2, σ_2^y, with unitary evolution for time t by the operator U:

$$\langle \sigma_1^x(t_0) \sigma_2^y(t_0+t) \rangle = \text{Tr}_{sys}[\rho_{sys}(t_0)\sigma_1^x U^\dagger \sigma_2^y U]$$
$$= \text{Tr}[U\rho_{sys}(t_0)\sigma_1^x U^\dagger \sigma_2^y]. \tag{C1}$$

This can be generalized to nonunitary channels [31],

$$\langle \sigma_1^x(t_0)\sigma_2^y(t_0+t) \rangle = \text{Tr}_{sys}[\sum_i K_i \rho_{sys}(t_0)\sigma_1^x K_i^\dagger \sigma_2^y]. \tag{C2}$$

To motivate this generalization, we consider Stinespring dilation. We can represent a general completely positive trace-preserving map by unitary evolution V acting on a combined system and bath, after which we trace out the bath

$$\rho_{sys} \to \sum_i K_i \rho_{sys} K_i^\dagger = \text{Tr}_{Bath}[V\rho_{bath}\rho_{sys}V^\dagger], \tag{C3}$$

where we do not just get ρ_{sys} back because the partial trace over operators on the composite space is in general not cyclic. With the addition of the bath, it is obvious that an analogous expression to Eq. 1 is the right one for the correlator

$$\langle \sigma_1^x(t_0)\sigma_2^y(t_0+t) \rangle = \text{Tr}_{sys,bath}[\rho_{bath}\rho_{sys}\sigma_1^x V^\dagger \sigma_2^y V]$$
$$= \text{Tr}_{sys,bath}[V\rho_{bath}\rho_{sys}\sigma_1^x V^\dagger \sigma_2^y]$$
$$= \text{Tr}_{sys}[\sum_i K_i \rho_{sys}\sigma_1^x K_i^\dagger \sigma_2^y]. \tag{C4}$$

We know how to apply the channel to $\rho_{sys}\sigma_x^1$ via the Choi isomorphism since it works for any matrix, not just density matrices. For any matrix M we feed in the transpose

$$\sum_i \text{Tr}_{anc}[\sum_{j_1,j_2} M_{j_2 j_1}|j_1\rangle\langle j_2| \sum_{k,l}|k\rangle K_i|k\rangle\langle l|\langle l|K_i^\dagger]$$
$$= \sum_{k,l} M_{kl} K_i|k\rangle\langle l|K_i^\dagger$$
$$= K_i M K_i^\dagger \tag{C5}$$

and get out the channel applied to M.

Using process shadows to estimate this correlation function, the traces over the in register and out register factorize

for each measurement result

$$\langle \sigma_1^x(0)\sigma_2^y(t)\rangle$$
$$\approx \frac{d}{m}\sum_{n=1}^{m} \mathrm{Tr}_{out}\Big[\mathrm{Tr}_{in}\Big[\big(\rho_{in}\sigma_1^x\big)^{\top} \prod_{r\in\text{qubit labels}} \Big(3|b_n^{in,r}\rangle\langle b_{\mu,n}^{in,r}| - \mathbb{I}\Big)\Big(3|b_{\mu,n}^{out,r}\rangle\langle b_{\mu,n}^{out,r}| - \mathbb{I}\Big)\Big] \sigma_2^y\Big]$$
$$= \frac{d}{m}\sum_{n} \mathrm{Tr}_{in}\Big[\big(\rho_{in}\sigma_1^x\big)^{\top} \prod_{r\in\text{qubit labels}} \Big(3|b_{\mu,n}^{in,r}\rangle\langle b_{\mu,n}^{in,r}| - \mathbb{I}\Big)\Big]\mathrm{Tr}_{out}\Big[\prod_{r}\Big(3|b_{\mu,n}^{out,r}\rangle\langle b_{\mu,n}^{out,r}| - \mathbb{I}\Big)\sigma_2^y\Big]$$
$$= \frac{d}{m}\sum_{n} \mathrm{Tr}_{in}\Big[\big(\rho_{in}\sigma_1^x\big)^{\top} \prod_{r\in\text{qubit labels}} \Big(3|b_{\mu,n}^{in,r}\rangle\langle b_{\mu,n}^{in,r}| - \mathbb{I}\Big)\Big]3b_{\mu,n}^{out,2}\,\delta_{\mu,Y}, \tag{C6}$$

and the trace over the output register drastically simplifies because we are time evolving a single site Pauli. Ultimately, this formula says (through the Kronecker delta $\delta_{\mu,Y}$) to throw out every output bitstring that did not

measure site 2 along the y axis, and if the measurement was along the y axis to weight the corresponding input trace with sign of the result.

[1] N. Gisin and R. Thew, Quantum communication, Nat. Photonics **1**, 165 (2007).

[2] A. S. Holevo, Bounds for the quantity of information transmitted by quantum communication channel, Probl. Infrom. Transm. **9**, 177 (1973).

[3] D. Kafri and J. M. Taylor, A noise inequality for classical forces, arXiv e-prints , arXiv:1311.4558 (2013), arXiv:1311.4558 [quant-ph].

[4] S. Bose, A. Mazumdar, G. W. Morley, H. Ulbricht, M. Toroš, M. Paternostro, A. A. Geraci, P. F. Barker, M. S. Kim, and G. Milburn, Spin Entanglement Witness for Quantum Gravity, Phys. Rev. Lett. **119**, 240401 (2017), arXiv:1707.06050 [quant-ph].

[5] C. Marletto and V. Vedral, Gravitationally Induced Entanglement between Two Massive Particles is Sufficient Evidence of Quantum Effects in Gravity, Phys. Rev. Lett. **119**, 240402 (2017), arXiv:1707.06036 [quant-ph].

[6] D. Carney, P. C. E. Stamp, and J. M. Taylor, Tabletop experiments for quantum gravity: a user's manual, Class. Quant. Grav. **36**, 034001 (2019), arXiv:1807.11494 [quant-ph].

[7] D. Carney, H. Müller, and J. M. Taylor, Using an Atom Interferometer to Infer Gravitational Entanglement Generation, PRX Quantum **2**, 030330 (2021), arXiv:2101.11629 [quant-ph].

[8] D. Gross, Y.-K. Liu, S. T. Flammia, S. Becker, and J. Eisert, Quantum state tomography via compressed sensing, Physical Review Letters **105**, 150401 (2010), arXiv: 0909.3304.

[9] I. L. Chuang and M. A. Nielsen, Prescription for experimental determination of the dynamics of a quantum black box, Journal of Modern Optics **44**, 2455 (1997).

[10] M. Branderhorst, J. Nunn, I. Walmsley, and R. Kosut, Simplified quantum process tomography, New Journal of Physics **11**, 115010 (2009).

[11] A. Shabani, R. L. Kosut, M. Mohseni, H. Rabitz, M. A. Broome, M. P. Almeida, A. Fedrizzi, and A. G. White, Efficient measurement of quantum dynamics via compressive sensing, Physical review letters **106**, 100401 (2011).

[12] S. T. Flammia, D. Gross, Y.-K. Liu, and J. Eisert, Quantum tomography via compressed sensing: error bounds, sample complexity and efficient estimators, New Journal of Physics **14**, 095022 (2012), publisher: IOP Publishing.

[13] H.-Y. Huang, R. Kueng, G. Torlai, V. V. Albert, and J. Preskill, Provably efficient machine learning for quantum many-body problems, Science **377** (2022).

[14] S. Aaronson, Shadow tomography of quantum states, in *Proceedings of the 50th annual ACM SIGACT symposium on theory of computing* (2018) pp. 325–338.

[15] H.-Y. Huang, R. Kueng, and J. Preskill, Predicting many properties of a quantum system from very few measurements, Nature Physics **16**, 1050 (2020).

[16] D. Zhu, Z. P. Cian, C. Noel, A. Risinger, D. Biswas, L. Egan, Y. Zhu, A. M. Green, C. H. Alderete, N. H. Nguyen, Q. Wang, A. Maksymov, Y. Nam, M. Cetina, N. M. Linke, M. Hafezi, and C. Monroe, Cross-platform comparison of arbitrary quantum states, Nature Communications **13**, 6620 (2022), number: 1 Publisher: Nature Publishing Group.

[17] G. Struchalin, Y. A. Zagorovskii, E. Kovlakov, S. Straupe, and S. Kulik, Experimental Estimation of Quantum State Properties from Classical Shadows, PRX Quantum **2**, 010307 (2021).

[18] T. Zhang, J. Sun, X.-X. Fang, X. Zhang, X. Yuan, and H. Lu, Experimental quantum state measurement with classical shadows, arXiv:2106.10190 (2021).

[19] T. Surawy-Stepney, J. Kahn, R. Kueng, and M. Guta, Projected least-squares quantum process tomography, Quantum **6**, 844 (2022).

[20] A. Jamiołkowski, Linear transformations which preserve trace and positive semidefiniteness of operators, Reports on Mathematical Physics **3**, 275 (1972).

[21] M.-D. Choi, Completely positive linear maps on complex matrices, Linear Algebra and its Applications **10**, 285 (1975).

[22] D. W. Leung, Choi's proof as a recipe for quantum process tomography, Journal of Mathematical Physics **44**,

528 (2003).

[23] S. Pirandola, R. Laurenza, C. Lupo, and J. L. Pereira, Fundamental limits to quantum channel discrimination, npj Quantum Information **5**, 50 (2019).

[24] D. Harlow and P. Hayden, Quantum computation vs. firewalls, Journal of High Energy Physics **2013**, 85 (2013), arXiv:1301.4504 [hep-th].

[25] D. Carney, Newton, entanglement, and the graviton, Physical Review D **105**, 10.1103/physrevd.105.024029 (2022).

[26] J. A. Miszczak, Z. Puchała, and P. Gawron, QI package for analysis of quantum states and operations (2011), https://github.com/iitis/qi.

[27] H. Bansal, Entanglement Cost Mathematica (2015),

https://github.com/hardikbansal/EntanglementCostMathematica.

[28] R. Levy, D. Luo, and B. K. Clark, Classical shadows for quantum process tomography on near-term quantum computers, arXiv:2110.02965 (2021).

[29] J. Helsen, M. Ioannou, I. Roth, J. Kitzinger, E. Onorati, A. H. Werner, and J. Eisert, Estimating gate-set properties from random sequences (2021), arXiv:2110.13178 [quant-ph].

[30] D. Gross, F. Krahmer, and R. Kueng, A Partial Derandomization of PhaseLift Using Spherical Designs, J Fourier Anal Appl **21**, 229.

[31] H. Carmichael, *An open systems approach to quantum optics*, Lecture notes in physics No. 18 (Springer-Verlag, Berlin ; New York, 1993).