







Robust estimation of nonlinear properties of quantum processesYuqing Wang ¹, Guoding Liu ^{1,*}, Zhenhuan Liu ^{1,†}, Yifan Tang ², Xiongfeng Ma ^{1,‡} and Hao Dai ^{3,§}¹*Center for Quantum Information, Institute for Interdisciplinary Information Sciences, Tsinghua University, Beijing 100084, People's Republic of China*²*Department of Mathematics and Computer Science, Freie Universität Berlin, 14195 Berlin, Germany*³*Beijing Institute of Mathematical Sciences and Applications (BIMSA), Huairou District, Beijing 101408, People's Republic of China*

(Received 4 June 2024; accepted 26 August 2024; published 12 September 2024)

The accurate and robust estimation of quantum process properties is crucial for quantum information processing and many-body physics. Combining classical shadow tomography and randomized benchmarking, Helsen *et al.* introduced a method to estimate the linear properties of quantum processes. In this work, we focus on the estimation protocols of nonlinear process properties that are robust to state preparation and measurement errors. We introduce two protocols, both utilizing random gate sequences but employing different postprocessing methods, which make them suitable for measuring different nonlinear properties. The first protocol offers a robust and sound method to estimate the out-of-time-ordered correlation, as demonstrated numerically in an Ising model. The second protocol estimates unitarity, effectively characterizing the incoherence of quantum channels. We expect the two protocols to be useful tools for exploring quantum many-body physics and characterizing quantum processes.

DOI: [10.1103/PhysRevA.110.032415](https://doi.org/10.1103/PhysRevA.110.032415)**I. INTRODUCTION**

With the development of quantum technology, the ability to control large quantum systems enables us to simulate extensive quantum many-body systems and investigate associated phenomena. Quantum processes, central to quantum physics, are mathematically described as quantum channels, which are completely positive and trace-preserving maps. All properties of interest within quantum processes are functions of the quantum channel, such as the scrambling strength [1], the unitarity [2], and the similarity between two quantum processes [3]. While quantum process tomography [4] offers a straightforward approach to estimating all these properties, its complexity grows exponentially with the number of qubits, rendering it impractical even for small-scale quantum systems.

Fortunately, full knowledge of a quantum channel is not always necessary to estimate some specific properties. Estimating partial knowledge of quantum channels can significantly reduce sample complexity compared to full tomography. This concept is analogous to progress in quantum state learning [5], such as classical shadow tomography [6] and its variations [7–10], which utilize random measurements for the efficient estimation of state properties. The process of classical shadow involves applying random unitary evolutions to the target state, followed by computational basis measurements, allowing for the simultaneous estimation of multiple state properties. This approach has been adapted for

channel property estimation through the Choi-Jamiołkowski isomorphism [11]. Specifically, one inputs random states to the quantum channel and performs randomized measurements on output states, equivalent to performing randomized measurements on the Choi state of the target channel. Like state shadow tomography, this technique allows the simultaneous estimation of multiple quantum channel properties.

While the aforementioned method can efficiently estimate certain channel properties without exponential sample complexities, it lacks robustness against state preparation and measurement (SPAM) errors. In many practical systems, SPAM errors, particularly measurement errors, can be as significant as or even surpass quantum gate errors [12,13]. Therefore, accurately estimating channel properties necessitates mitigating the influence of SPAM errors. Randomized benchmarking serves as a widely adopted protocol for this purpose, allowing property estimation of quantum channels while reducing the impact of SPAM errors [14]. However, conventional randomized benchmarking protocols are limited in measurable properties, primarily restricted to properties such as average fidelity [15].

Utilizing group twirling and fitting techniques, Helsen *et al.* have combined randomized benchmarking with classical shadow to estimate arbitrary linear properties of quantum channels, robust against SPAM errors [16]. However, to fully explore quantum phenomena and characterize quantum channels, linear properties alone are insufficient. Many critical nonlinear properties exist, such as out-of-time-ordered correlation (OTOC) [1], a measure of information scrambling. The quantity of OTOC is critical in both quantum many-body physics [17,18] and quantum information [19,20]. Yet, current estimation methods for OTOC lack robustness against SPAM errors [21–27], which poses a challenge to observe

*Contact author: lgd22@mails.tsinghua.edu.cn†Contact author: liu-zh20@mails.tsinghua.edu.cn‡Contact author: xma@tsinghua.edu.cn§Contact author: dhao@bimsa.cn

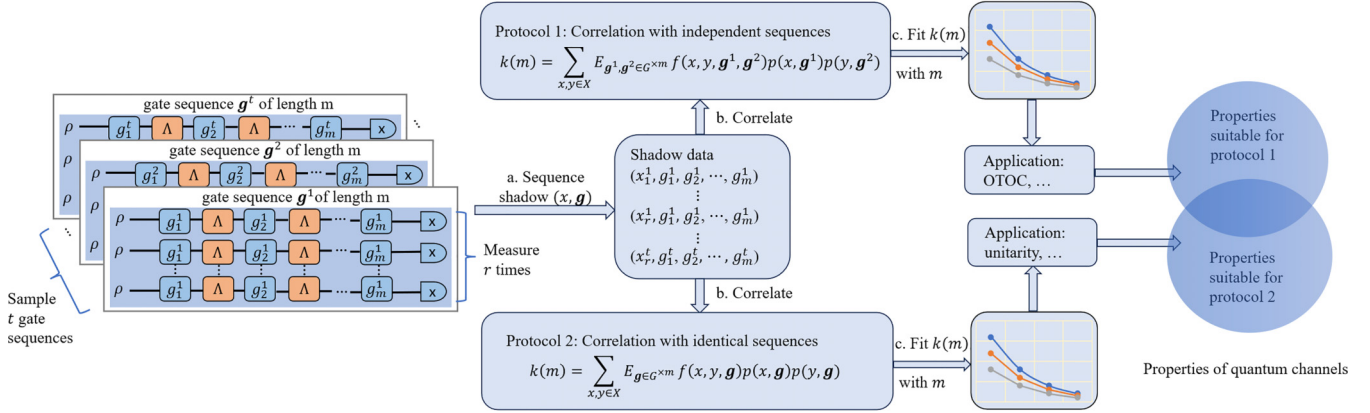


FIG. 1. The workflow of the nonlinear channel properties estimation via the generalized uniformly independent random sequence protocol introduced in Secs. III A and III B. In step “a,” we obtain the shadow sequence data (x, \mathbf{g}) by implementing random gate sequences and POVM measurements. For each sequence length m , we measure the same gate sequences for r times and iterate the process with t different gate sequences. In step “b,” we calculate the correlation function $k(m)$ using the shadow sequence data. The first protocol utilizes data from independent sequences, while the second protocol involves data from identical sequences. The entire procedure is repeated for several different sequence lengths m . In step “c,” we perform a fitting of the correlation function $k(m)$ against the sequence length m to extract the desired nonlinear channel properties. With different group G , the two protocols may be suitable for evaluating different channel properties. In particular, if G is the global n -qubit Clifford group, protocol 1 can estimate all the properties that protocol 2 can obtain, as discussed in Sec. III C.

information scrambling. In addition, other nonlinear properties, such as unitarity and nonstabilizerness of a quantum channel, are also important and widely discussed in the quantum information field [2,28].

Building upon Helsen *et al.*'s framework [16], we introduce two protocols to estimate nonlinear properties of quantum channels with robustness against SPAM errors, as shown in Fig. 1. Both protocols implement random gate sequences sampled from a group and perform a positive operator-valued measure (POVM) to collect shadow data in the first step. Then, they differ in classical data postprocessing. The first protocol utilizes correlations between measurement data from independently chosen gate sequences, while the second harnesses correlations from identical gate sequences. The expectation of the correlation is a multiple exponential decay function with the circuit depth. Through exponential fitting, one can get nonlinear properties of quantum channels excluding the SPAM error. Different classical postprocessing procedures make the two protocols suitable for evaluating different channel properties and exploring different phenomena within quantum many-body systems. As an application, we employ the first protocol to measure OTOC and theoretically analyze its sample complexity. In a long-range interaction Ising model, we numerically demonstrate the effectiveness of the protocol and resilience to SPAM errors. Moreover, we explore the potential of the second protocol in estimating unitarity [2]. We analyze the type of channel properties that can be measured with each protocol when the matrix exponential fitting is allowed. We find that when the gate sequences are sampled from a unitary two-design group, such as the Clifford group, the measurable quantities of the first protocol cover those of the second one.

This work is organized as follows. In Sec. II, we introduce basic notations and the necessary preliminaries for this work. In Sec. III, we present two protocols to estimate nonlinear channel properties, provide two illustrative examples, and discuss the measurable properties of these two protocols. In

Sec. IV, we showcase the results of our simulations pertaining to OTOC estimation. We conclude in Sec. V.

II. PRELIMINARIES

In this section, we define key notations and review essential concepts in the Pauli-Liouville representation, the Clifford group, and the uniformly independent random sequence (UIRS) shadow [16].

A. Pauli-Liouville representation

For an n -qubit system, \mathbb{C}^d with the dimension $d = 2^n$, the normalized Pauli group is defined as

$$\mathbb{P}_n = \left\{ \frac{\sigma}{\sqrt{d}} \mid \sigma \in \{I, \sigma_x, \sigma_y, \sigma_z\}^{\otimes n} \right\}. \quad (1)$$

The elements of this group, I, σ_x, σ_y , and σ_z , represent the identity operator, and Pauli X, Y , and Z operators, respectively, defined as

$$I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \\ \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \quad (2)$$

In certain contexts, the identity operator is excluded, leading to the set $\mathbb{P}_n^0 = \mathbb{P}_n \setminus \frac{1}{\sqrt{d}}$, where we define $\mathbf{1} = I^{\otimes n}$. The Pauli group forms an orthonormal basis in the linear operator space $L(\mathbb{C}^d)$, equipped with the inner product $\langle A, B \rangle = \text{Tr}(A^\dagger B)$. Consequently, any operator O can be expressed as a sum over this basis,

$$O = \sum_{\sigma_i \in \mathbb{P}_n} \langle \sigma_i, O \rangle \sigma_i. \quad (3)$$

The Pauli-Liouville representation vectorizes the operator space, defined as the linear map $|\cdot\rangle\rangle : L(\mathbb{C}^d) \rightarrow \mathbb{C}^{d^2}$, by assigning $|\sigma_i\rangle\rangle = e_i$, where $\sigma_i \in \mathbb{P}_n$ and $\{e_i\}$ forms an orthonormal basis in \mathbb{C}^{d^2} . Furthermore, the inner product in the Hilbert space \mathbb{C}^{d^2} is represented as $\langle\langle A|B\rangle\rangle = \langle A, B \rangle = \text{Tr}(A^\dagger B)$. By linearity, the Pauli-Liouville representation of an operator O is expressed as

$$|O\rangle\rangle = \sum_{\sigma_i \in \mathbb{P}_n} \langle\sigma_i, O\rangle |\sigma_i\rangle\rangle. \quad (4)$$

For a quantum channel $\mathcal{E} : L(\mathbb{C}^d) \rightarrow L(\mathbb{C}^d)$, it can be represented as a matrix in the Pauli-Liouville representation. The elements of this matrix are given by

$$\mathcal{E}_{i,j} = \langle\langle \sigma_i | \mathcal{E} | \sigma_j \rangle\rangle = \langle\langle \sigma_i | \mathcal{E}(\sigma_j) \rangle\rangle = \text{Tr}[\sigma_i \mathcal{E}(\sigma_j)]. \quad (5)$$

In this representation, the composition of channels corresponds to matrix multiplication. Specifically, for two quantum channels \mathcal{E}_1 and \mathcal{E}_2 acting on an operator O , the composition is expressed as

$$|\mathcal{E}_1 \circ \mathcal{E}_2(O)\rangle\rangle = \mathcal{E}_1 \mathcal{E}_2 |O\rangle\rangle. \quad (6)$$

B. Clifford group

The Clifford group \mathcal{C}_n comprises n -qubit unitary operators that normalize the Pauli group \mathbb{P}_n . An operator U is an element of the Clifford group if and only if, for every Pauli operator $P \in \mathbb{P}_n$, the conjugated operator UPU^\dagger is also a Pauli operator, up to a global phase. Mathematically, this is expressed as follows: for each $P \in \mathbb{P}_n$ and $U \in \mathcal{C}_n$, there exists $P' \in \mathbb{P}_n$ and a global phase $e^{i\theta\pi}$ such that $UPU^\dagger = e^{i\theta\pi} P'$.

In the Pauli-Liouville representation, the Clifford group decomposes into two nonequivalent irreducible representations, $\forall g \in \mathcal{C}_n$,

$$\omega(g) = \tau_{\text{tr}}(g) \oplus \tau_{\text{ad}}(g), \quad (7)$$

where τ_{tr} is the trivial representation supported on the normalized identity matrix, $|\mathbf{1}/\sqrt{d}\rangle\rangle$, and τ_{ad} is the adjoint representation supported on the traceless matrices. The projector onto the representation space of τ_{tr} is $P_{\text{tr}} = |\mathbf{1}\rangle\rangle\langle\langle \mathbf{1}|/d$, while the projector onto the irreducible representation space of τ_{ad} is given by $P_{\text{ad}} = \sum_{\sigma \in \mathbb{P}_n} |\sigma\rangle\rangle\langle\langle \sigma|$.

C. Uniformly independent random sequence shadow

The uniformly independent random sequence (UIRS) shadow [16] can estimate the linear properties of noisy gate sets by combining the methods of randomized benchmarking and classical shadow tomography. The protocol starts with the collection of shadow sequence data obtained from a random sequence of gates, followed by a POVM measurement. The subsequent step is computing the correlation function from the shadow data, which encodes the desired properties of the gate set. The final step entails estimating the gate set properties through a fitting process [16]. Notably, the UIRS protocol exhibits robustness against SPAM errors, similar to the conventional randomized benchmarking protocol.

Now, we proceed to formalize the UIRS protocol mathematically, following the procedures depicted in Fig. 1. Let $\mathbf{g} = (g_1, \dots, g_m)$ denote a random sequence, comprising m

gates where each gate g_i is uniformly and independently selected from a unitary gate set \mathbb{G} . Typically, \mathbb{G} is chosen as a group and we adhere to this convention in our work. Given an input state ρ and a POVM with a finite-outcomes set \mathcal{X} , $\{E_x\}_{x \in \mathcal{X}}$, each iteration of the protocol yields a piece of shadow data, denoted as (x, \mathbf{g}) . These data contain the measurement outcome x and the random gate sequence \mathbf{g} . The probability of acquiring this specific data is given by

$$p(x, \mathbf{g}) = \langle\langle E_x | \mathcal{E}(\mathbf{g}) | \rho \rangle\rangle, \quad (8)$$

where $\mathcal{E}(\mathbf{g})$ denotes the noisy implementation of the random sequence \mathbf{g} .

In experiments, the presence of noise is inevitable. We model the noisy initial state and measurement as $\tilde{\rho}$ and $\{\tilde{E}_x\}$, respectively. Regarding the noisy implementation of the random sequence, we assume gate-independent noise, a special type of Markovian noise. Specifically, for each gate g_i in the random sequence, the ideal unitary implementation is given by

$$\omega(g_i) | \rho \rangle\rangle = | g_i(\rho) \rangle\rangle, \quad (9)$$

where ω is the Liouville representation of the group \mathbb{G} and $g_i(\rho) = U_{g_i} \rho U_{g_i}^\dagger$. The actual implementation of the noisy gate is given by

$$\phi(g_i) = \Lambda_L \omega(g_i) \Lambda_R, \quad (10)$$

where Λ_L and Λ_R are noise channels independent of the gate choice. Then, the channel $\Lambda = \Lambda_R \Lambda_L$ denotes the noise occurring between gates. Consequently, the overall implemented channel for the random sequence is

$$\mathcal{E}(\mathbf{g}) = \prod_{i=1}^m \phi(g_i). \quad (11)$$

Under these assumptions and considering errors in state and measurement, the probability of obtaining data (x, \mathbf{g}) is given by

$$p(x, \mathbf{g}) = \langle\langle \tilde{E}_x | \prod_{i=1}^m \phi(g_i) | \tilde{\rho} \rangle\rangle. \quad (12)$$

After multiple independent experimental rounds, we acquire a collection of gate-set shadows, $\{(x_i, \mathbf{g}_i)\}_{i=1}^S$. From the shadow data, we aim to extract meaningful information, typically in the form of an expectation value of a sequence correlation function. We consider a correlation function $f(x, \mathbf{g}, m) : \mathcal{X} \times \mathbb{G}^{\times m} \times \mathbb{Z}^+ \rightarrow \mathbb{C}$, with a specific form,

$$f_A(x, \mathbf{g}, m) = \text{Tr} \left[B_x \tau(g_m) \prod_{i=1}^{m-1} A \tau(g_i) \right], \quad (13)$$

where τ represents an irreducible representation of group \mathbb{G} , and A and B_x denote preselected operators supported on the representation space associated with τ . The expectation of this correlation function across all possible choices of random gate sequences is given by

$$k_A(m) = \mathbb{E}_{\mathbf{g} \in \mathbb{G}^{\times m}} \sum_{x \in \mathcal{X}} f_A(x, \mathbf{g}, m) p(x, \mathbf{g}). \quad (14)$$

The quantity $k_A(m)$ is always referred to as the correlator with sequence correlation function f_A . It has been proven that

$k_A(m)$ has an exponential behavior in m [16],

$$k_A(m) = \text{Tr}\{\Theta(\{E_x\}_x, \rho)[\Phi(A, \Lambda)]^{m-1}\}, \quad (15)$$

where $\Theta(\{E_x\}_x, \rho)$ is a matrix dependent solely on SPAM and B_x ; $\Phi(A, \Lambda)$ is a matrix dependent on \mathbb{G} , A , and Λ . If the Liouville representation of \mathbb{G} has decomposition, $\omega(g) = \tau(g)^{\text{tr}} \oplus \omega'(g)$, where $\omega'(g)$ contains no copy of $\tau(g)$, then $\Phi(A, \Lambda)$ reduces to

$$\Phi(A, \Lambda)_{i,j} = \frac{1}{|P_j|} \text{Tr}(P_j A P_j \Lambda), \quad (16)$$

where P_i is the projector onto the i th copy of $\tau(g)$ inside the representation of $\omega(g)$ and $|P_j| = \text{rank}(P_j)$ represents the dimension of $\tau(g)$. With the exponential fitting, one can obtain $\Phi(A, \Lambda)$ while excluding the influence of the SPAM, ρ and E_x . Each matrix element of $\Phi(A, \Lambda)$, $\text{Tr}(P_i A P_j \Lambda)$, is a linear function of Λ dependent on A . We can thus estimate the linear properties of Λ .

In reality, we cannot get the correlator k_A perfectly due to the finite sampling, but we can construct an estimator, $\hat{k}_A(m)$, from the shadow data such that $\hat{k}_A(m)$ converges to $k_A(m)$ when the number of experiment rounds is sufficiently large. A standard estimator is given by

$$\hat{k}_A(m) = \frac{1}{S} \sum_{i=1}^S f_A(x_i, \mathbf{g}_i, m). \quad (17)$$

When $S \rightarrow \infty$, this estimator will converge to $k_A(m)$ and allow one to get $\Phi(A, \Lambda)$.

In summary, the UIRS protocol serves as a valuable tool for estimating linear properties of noise channels associated with gate sets. Alternatively, when considering a noiseless gate set \mathbb{G} and artificially inserting the noise channel Λ between two consecutive random gates g_i and g_{i+1} , the UIRS protocol can be interpreted as a method for extracting properties of any given channel. Typically, Λ can be chosen as a unitary evolution, which is of particular interest in studies of quantum many-body systems. In the following section, we will extend the UIRS protocol to the estimation of nonlinear properties of quantum channels, focusing on evaluating the OTOC of unitary quantum evolution and the unitarity of the quantum channels.

III. NONLINEAR CHANNEL PROPERTIES ESTIMATION VIA GENERALIZED UIRS

In this section, we present the generalized UIRS protocol to estimate nonlinear channel properties. Our protocol shares the quantum procedure with the UIRS protocol, that is, applying the random gate sequences and measurements to obtain the shadow data (x, \mathbf{g}) . The difference lies in the classical postprocessing stage. In the UIRS protocol, the correlation function only involves one piece of shadow data, such as Eq. (13). In our protocol, we evaluate the correlation function by incorporating two pieces of shadow data instead of one. The gate sequences in two pieces of shadow data can be chosen as independent or identical, corresponding to two different protocols. We study the two cases separately and draw a comparative analysis.

A. Nonlinear correlation via independent sequences

We first introduce the generalized UIRS protocol utilizing two pieces of shadow data derived from independent gate sequences. To illustrate, we provide a specific example demonstrating the effective estimation of the OTOC within this protocol.

After obtaining the shadow data, we can construct a sequence correlation function that encodes the information of the desired quantity. In this context, we consider a second-order sequence correlation function defined as

$$f(x, y, \mathbf{g}^1, \mathbf{g}^2, m) : \mathcal{X} \times \mathcal{X} \times \mathbb{G}^{\times m} \times \mathbb{G}^{\times m} \times \mathbb{Z}^+ \rightarrow \mathbb{C}, \quad (18)$$

which depends on two independent random sequences \mathbf{g}^1 and \mathbf{g}^2 and their measurement outcomes x and y . Notice that the function is analogous to the observable with respect to two copies of the state in the case of state shadow tomography [6]. The expectation of the correlation function is taken over the group $\mathbb{G}^{\times m} \times \mathbb{G}^{\times m}$ and the measurement outcomes,

$$k_f(m) = \sum_{x, y \in \mathcal{X}} \mathbb{E}_{\mathbf{g}^1, \mathbf{g}^2 \in \mathbb{G}^{\times m}} f(x, y, \mathbf{g}^1, \mathbf{g}^2, m) p(x, \mathbf{g}^1) p(y, \mathbf{g}^2). \quad (19)$$

With the sequence shadow data $\{(x_i, \mathbf{g}^i)\}_{i=1}^S$, we can construct an estimator $\hat{k}_f(m)$ for $k_f(m)$ as shown below,

$$\hat{k}_f(m) = \frac{1}{S(S-1)} \sum_{i \neq j} f(x_i, x_j, \mathbf{g}^i, \mathbf{g}^j, m). \quad (20)$$

When the number of samples S tends to infinity, the estimator $\hat{k}_f(m)$ converges to $k_f(m)$.

In this work, we investigate a special kind of correlation function in the form of

$$\begin{aligned} & f_A(x, y, \mathbf{g}^1, \mathbf{g}^2, m) \\ &= \text{Tr} \left\{ B_{xy} [\tau_1(\mathbf{g}_m^1) \otimes \tau_2(\mathbf{g}_m^2)] \prod_{i=1}^{m-1} A [\tau_1(\mathbf{g}_i^1) \otimes \tau_2(\mathbf{g}_i^2)] \right\}, \end{aligned} \quad (21)$$

where τ_1 and τ_2 are both irreducible representations of group \mathbb{G} , and A, B_{xy} are operators supported on the corresponding representation space of $\tau_1 \otimes \tau_2$. In this case, the expectation value $k_f(m)$ is denoted as $k_A(m)$ since it encodes information of the operator A . For the special-type correlation function of Eq. (21), Theorem 1 tells us that the expectation value $k_A(m)$ exhibits an exponential decay behavior with respect to the sequence length m . The complete proof is available in Appendix A.

Theorem 1. (Exponential decay). Given a generalized UIRS protocol with group \mathbb{G} , the expectation value $k_A(m)$ of the second-order correlation function given by Eq. (21) has the following form:

$$k_A(m) = \text{Tr}\{\Theta(\{E_x\}, \rho)[\Phi(A, \Lambda)]^{m-1}\}, \quad (22)$$

where $\Theta(\{E_x\}, \rho)$ and $\Phi(A, \Lambda)$ are induced matrices dependent on SPAM and channel Λ , respectively. If $\omega(g) = \tau_1(g)^{\text{tr}} \oplus \omega'_1(g)$ where $\omega'_1(g)$ contains no copy of $\tau_1(g)$ and

$\omega(g) = \tau_2(g)^{n\tau_2} \oplus \omega'_2(g)$ where $\omega'_2(g)$ contains no copy of $\tau_2(g)$, then the matrix $\Phi(A, \Lambda)$ has the element

$$[\Phi(A, \Lambda)]_{i\nu, j\nu'} = \frac{1}{|P_j||P_{j'}|} \text{Tr}[(P_j \otimes P_{j'})^A (P_i \otimes P_{i'}) \Lambda^{\otimes 2}], \quad (23)$$

where P_i and P_j are the projectors onto the i th and j th copies of τ_1 inside the representation ω , respectively, and $P_{i'}$ and $P_{j'}$ are the projectors onto the i' th and j' th copies of τ_2 inside the representation ω , respectively. In the special case that the representation ω only has one copy of τ_1 and one copy of τ_2 , the matrix $\Phi(A, \Lambda)$ reduces to a number.

As an application of this generalized UIRS protocol, we demonstrate that OTOC can be estimated via the generalized UIRS protocol robust to SPAM errors. We first briefly review the definition of OTOC, which originates from information scrambling. Information scrambling is one of the signatures of quantum chaos in many-body systems, and the strength of the scrambling can be quantitatively characterized by OTOC [1, 29]. Specifically, for a local operator V , after a unitary dynamics $U_t = e^{-iHt}$ controlled by a Hamiltonian H , the local information encoded by V can spread over many sites and become nonlocal. Assume that there is another operator W located in a different position from V and define $V(t) = U_t^\dagger V U_t$. Initially, $V = V(0)$ is localized and commutes with W , but $V(t)$ will be nonlocal and become noncommutative with W when $V(t)$ is spread to the position of W . Consequently, the operator growth associated with U_t can be characterized by the expectation value of the squared commutator of $V(t)$ and W over a state ρ ,

$$C(t) = \langle [V(t), W]^\dagger [V(t), W] \rangle_\rho. \quad (24)$$

A closely related quantity is OTOC, defined by

$$O(t) = \langle W^\dagger V(t)^\dagger W V(t) \rangle_\rho, \quad (25)$$

which satisfies

$$C(t) = 2\{1 - \text{Re}[O(t)]\}. \quad (26)$$

The quantities $O(t)$ and $C(t)$ both measure the strength of information scrambling. Here, we take ρ as the maximally mixed state, which can also be regarded as the thermal state of infinite temperature. In this case, OTOC has an explicit form,

$$O(t) = \frac{1}{d} \text{Tr}[W^\dagger V(t)^\dagger W V(t)]. \quad (27)$$

Below, we focus on the multiqubit system with dimension $d = 2^n$. We consider V and W to be nontrivial and unnormalized Pauli operators, $V, W \in \sqrt{d} \mathbb{P}_n^0$. The situation can also be straightforwardly generalized to the case that V and W are arbitrary operators, as shown in Appendix B. In the following theoretical analysis, we assume that the random gate is taken from the multiqubit Clifford group and is noiseless. To robustly estimate the OTOC against SPAM errors, we introduce an adjustment of the generalized UIRS protocol by inserting a unitary gate $\mathcal{U}_t(\rho) = U_t \rho U_t^\dagger$ between two random Clifford gates. This can be interpreted as considering the unitary evolution U_t as the noise, as depicted in Fig. 2. In reality, the Clifford gates $\{g_i\}$ are also noisy and we can regard their noises as being absorbed by the unitary gate. For instance, when considering $\phi(g_i) = \Lambda_L \omega(g_i) \Lambda_R$ such as Eq. (10), we

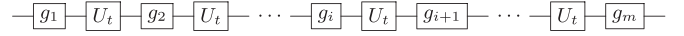


FIG. 2. Random Clifford gate sequence g_1, g_2, \dots, g_m intertwined with a fixed unitary gate, U_t .

can view the inserted gate to be a noisy implementation of the unitary gate, $\tilde{U}_t = \Lambda_R U_t \Lambda_L$, which allows us to obtain the information of the noisy unitary dynamics robust to SPAM errors. In this sense, the implemented Clifford gate is ideal, and then our protocol can estimate the OTOC of the noisy unitary evolution without the influence of Clifford gate noises. Meanwhile, in practice, one can use an approximate 2-design gate set to replace the Clifford group in this protocol without compromising the faithfulness of the estimation, which we will detail in Appendix C. Since an n -qubit approximate 2-design gate set can be implemented in a one-dimensional $O(\ln n)$ -depth Clifford circuit [30], our protocol of OTOC estimation is also implementable in a short-depth circuit with restricted connections.

For estimating OTOC, we define the second-order correlation function as

$$f_A(x, y, \mathbf{g}^1, \mathbf{g}^2, m) = \text{Tr} \left[B_{xy} \tau_{\text{ad}}(g_m^1) \otimes \tau_{\text{ad}}(g_m^2) \prod_{i=1}^{m-1} A \tau_{\text{ad}}(g_i^1) \otimes \tau_{\text{ad}}(g_i^2) \right], \quad (28)$$

where $A = (d^2 - 1)^2 \sum_{\sigma \in \mathbb{P}_n^0} \text{Tr}(W \sigma W \sigma) |\sigma \otimes \sigma\rangle \langle V \otimes V|$, $B_{xy} = |\rho\rangle \langle E_x| \otimes |\rho\rangle \langle E_y|$, and τ_{ad} is the adjoint representation supported on the traceless matrices. After obtaining shadow sequence data, we can estimate the expectation value $k_A(m)$. Next, we show the relationship between $k_A(m)$ and OTOC.

Corollary 1. The expectation value $k_A(m)$ of the second-order correlation function defined in Eq. (28) has an exponential decay behavior with respect to the sequence length m , $k_A(m) = a[dO(t)]^{m-1}$. The decay parameter is $p(A) = dO(t)$ with $O(t) = \frac{1}{d} \text{Tr}[W^\dagger V(t)^\dagger W V(t)]$ being OTOC.

The full proof of Corollary 1 is available in Appendix B. Thus, OTOC can be robustly evaluated with the protocol proposed in this section, with the simulation shown in Sec. IV.

To analyze the sample complexity of this protocol, we evaluate the variance of the correlation function defined in Eq. (28). An upper bound is given in the following theorem and the full proof of the variance is detailed in Appendix D.

Theorem 2. Viewing f_A defined in Eq. (28) as a random variable with probability distribution from the shadow data, $p(x, \mathbf{g}^1)p(y, \mathbf{g}^2)$, the variance of this random variable is upper bounded by $O(d^{8m-12})$.

In the experiment, we obtain S shadow data samples to estimate the expectation value $k_A(m)$ through Eq. (20) and then derive the OTOC. The variance of the estimator $\hat{k}_A(m)$ depends on both S and the variance of the random variable f_A with probability distribution $p(x, \mathbf{g}^1)p(y, \mathbf{g}^2)$. As a corollary of Theorem 2, the variance of the estimator $\hat{k}_A(m)$ also has an upper bound.

Corollary 2. Suppose the number of samples is S and the estimator for $k_A(m, S)$ is

$$\hat{k}_A(m, S) = \frac{1}{S(S-1)} \sum_{i \neq j} f(x_i, x_j, \mathbf{g}^i, \mathbf{g}^j, m). \quad (29)$$

Then the variance has an upper bound,

$$\begin{aligned} & \mathbf{Var}[\hat{k}_A(m, S)] \\ & \leq \begin{cases} \frac{1}{S(S-1)} O(d^{-4}) + \frac{2(S-2)}{S(S-1)} O(d^{-1}), & m = 1 \\ \frac{1}{S(S-1)} O(d^4) + \frac{2(S-2)}{S(S-1)} O(1), & m = 2 \\ \frac{1}{S(S-1)} O(d^{8m-12}) + \frac{2(S-2)}{S(S-1)} O(d^{4m-4}), & m > 2. \end{cases} \end{aligned} \quad (30)$$

Particularly, we can estimate OTOC by the ratio $\hat{k}_A(2, S)/\hat{k}_A(1, S)$ with S samples. From Corollary 2, the variance of the estimators $\hat{k}_A(2, S)$ and $\hat{k}_A(1, S)$ can be upper bounded by $O(d^4 S^{-2}) + O(S^{-1})$ and $O(d^{-4} S^{-2}) + O(d^{-1} S^{-1})$, respectively. Then, the variance of the ratio has an upper bound given by

$$\begin{aligned} & \mathbf{Var}\left(\frac{\hat{k}_A(2, S)}{\hat{k}_A(1, S)}\right) \\ & \leq \left(\frac{\mathbb{E}[\hat{k}_A(2, S)]}{\mathbb{E}[\hat{k}_A(1, S)]}\right)^2 \left[\frac{\mathbf{Var}[\hat{k}_A(2, S)]}{\{\mathbb{E}[\hat{k}_A(2, S)]\}^2} + \frac{\mathbf{Var}[\hat{k}_A(1, S)]}{\{\mathbb{E}[\hat{k}_A(1, S)]\}^2} \right] \\ & = O\left(\frac{d^8}{S^2}\right) + O\left(\frac{d^5}{S}\right), \end{aligned} \quad (31)$$

where the two expectation values can be evaluated by $\mathbb{E}[\hat{k}_A(2, S)] = O(d^{-1})$ and $\mathbb{E}[\hat{k}_A(1, S)] = O(d^{-2})$. The detailed calculation of the variances and expectations can be found in Appendix D. Thus, the number of samples S only needs to take $O(\frac{d^5}{\epsilon})$ to evaluate the estimate within precision ϵ .

B. Nonlinear correlation via identical sequences

The nonlinear correlation function defined in Eq. (18) is connected with two independent pieces of sequences. In this part, we set the two gate sequences to be identical and study the corresponding second-order correlation function, where we find that the protocol estimating unitarity robustly presented in Ref. [2] is a specific instance of this protocol.

Specifically, we consider the second-order correlation function when the two random gate sequences are chosen to be the same, i.e., $\mathbf{g}^1 = \mathbf{g}^2$, defining

$$k_f(m) = \sum_{x, y \in \mathcal{X}} \mathbb{E}_{\mathbf{g} \in \mathbb{G}^{\otimes m}} f(x, y, \mathbf{g}, m) p(x, \mathbf{g}) p(y, \mathbf{g}). \quad (32)$$

Similarly to the case for independent sequences, we assume

$$f(x, y, \mathbf{g}, m) = f_A(x, y, \mathbf{g}, m) = \text{Tr} \left[B_{xy} \tau(g_m) \prod_{i=1}^{m-1} A \tau(g_i) \right], \quad (33)$$

where τ is an irreducible representation of the group \mathbb{G} , and A, B_{xy} are matrices supported on the representation space

associated with τ . The following theorem states that the expectation value $k_A(m) = k_{f_A}(m)$ has an exponential decay, with the proof shown in Appendix E.

Theorem 3. (Exponential decay). Given a generalized UIRS protocol with respect to the group \mathbb{G} , the expectation value $k_A(m)$ defined with Eqs. (32) and (33) has the following form:

$$k_A(m) = \text{Tr} \{ \Theta(\{E_x\}, \rho) [\Phi(A, \Lambda)]^{m-1} \}, \quad (34)$$

where $\Theta(\{E_x\}, \rho)$ and $\Phi(A, \Lambda)$ are induced matrices dependent on SPAM and channel Λ , respectively. The matrix $\Phi(A, \Lambda) = P_\tau (A \otimes \Lambda^{\otimes 2}) P_\tau$ has the element

$$[\Phi(A, \Lambda)]_{i,j} = \frac{1}{|P_j|} \text{Tr}(P_j A^T P_i \Lambda^{\otimes 2}), \quad (35)$$

where $P_\tau = \mathbb{E}_{g \in \mathbb{G}} \tau(g) \otimes \omega(g)^{\otimes 2}$, and P_i and P_j are the projectors onto the i th and j th copies of τ inside $\omega^{\otimes 2}$, respectively. In the special case that the representation $\omega^{\otimes 2}$ only has one copy of τ , the matrix $\Phi(A, \Lambda)$ reduces into a number.

Below, we show that the protocol in this section can be used to estimate unitarity. For the between-gates noise channel $\Lambda : \mathcal{B}(\mathbb{C}^d) \rightarrow \mathcal{B}(\mathbb{C}^d)$, the unitarity of the channel is defined as the average purity of the output states with the identity part being subtracted,

$$u(\Lambda) = \frac{d}{d-1} \int d\psi \text{Tr}[\Lambda'(|\psi\rangle\langle\psi|)^\dagger \Lambda'(|\psi\rangle\langle\psi|)], \quad (36)$$

where $\Lambda'(A) = \Lambda(A) - \frac{\text{Tr}\Lambda(A)}{d} \mathbf{1}$.

The unitarity can be efficiently and robustly estimated against SPAM errors through an experimental protocol based on randomized benchmarking [2]. Here, we prove that unitarity can be estimated under the generalized UIRS protocol when the two involved random sequences are identical and the group \mathbb{G} is the Clifford group. Again, we assume the noise channel of the Clifford gates to be gate independent.

Given a state ρ , a sequence of gates, $\{g_1, g_2, \dots, g_m\}$, and an observable $E = \sum_x E_x$ with $\{E_x\}_x$ representing a computational-basis measurement, it has been proven that the square of the measurement result has an exponential decay after taking an expectation over the Clifford group \mathbb{G} [2],

$$\mathbb{E}_{\mathbf{g} \in \mathbb{G}^{\otimes m}} \left(\text{Tr} \left\{ \tilde{E} \prod_{i=1}^m \phi(g_i) [\tilde{\rho}] \right\} \right)^2 = a + bu(\Lambda)^{m-1}, \quad (37)$$

where a and b are fitting constants and $u(\Lambda)$ is the unitarity of the noise channel. The notation $\tilde{\cdot}$ represents the noisy versions of quantum states, gates, and observables.

We demonstrate that Eq. (37) can be written in the form of Eq. (32) by properly selecting the operator A and B_{xy} . Let $A = \mathbf{1}$ and $B_{xy} = \mathbf{1}$, and set the irreducible representation τ as the trivial representation of the Clifford group. Then, the second-order correlation function is

$$f_A(x, y, \mathbf{g}, m) = \text{Tr} \left[\mathbf{1} \prod_{i=1}^m \tau_{\text{tr}}(g_i) \right] = 1. \quad (38)$$

Therefore,

$$\begin{aligned}
k_A(m) &= \mathbb{E}_{\mathbf{g} \in \mathbb{G}^{\times m}} \sum_{x,y} f_A(x,y,\mathbf{g},m) p(x,\mathbf{g}) p(y,\mathbf{g}) \\
&= \mathbb{E}_{\mathbf{g} \in \mathbb{G}^{\times m}} \sum_{x,y} \text{Tr} \left\{ \tilde{E}(x) \prod_{i=1}^m \phi(g_i)[\tilde{\rho}] \right\} \\
&\quad \times \text{Tr} \left\{ \tilde{E}(y) \prod_{i=1}^m \phi(g_i)[\tilde{\rho}] \right\} \\
&= \mathbb{E}_{\mathbf{g} \in \mathbb{G}^{\times m}} \left[\text{Tr} \left\{ \tilde{E} \prod_{i=1}^m \phi(g_i)[\tilde{\rho}] \right\} \right]^2. \tag{39}
\end{aligned}$$

According to Theorem 3, the quantity $k_A(m)$ has an exponential decay with respect to the matrix $\Phi(A, \Lambda) = P_\tau(A \otimes \Lambda^{\otimes 2})P_\tau$, where

$$P_\tau = \mathbb{E}_{g \in \mathbb{G}} \tau_{\text{tr}}(g) \otimes \omega(g)^{\otimes 2} \tag{40}$$

is the projector onto the trivial subspace within $\omega^{\otimes 2}$. The explicit form of the projector is

$$P_\tau = |B_1 \otimes B_1\rangle\langle B_1 \otimes B_1| + |B_1 \otimes B_2\rangle\langle B_1 \otimes B_2|, \tag{41}$$

where the two operators are

$$B_1 = \frac{\mathbf{1}_{d^2}}{d}, \quad B_2 = \frac{F - B_1}{\sqrt{d^2 - 1}}. \tag{42}$$

Here, F is a SWAP operator on the space of $\Lambda^{\otimes 2}$. When Λ is trace preserving,

$$\begin{aligned}
\Phi(A, \Lambda) &= \begin{pmatrix} \langle\langle B_1 | \Lambda^{\otimes 2} | B_1 \rangle\rangle & \langle\langle B_1 | \Lambda^{\otimes 2} | B_2 \rangle\rangle \\ \langle\langle B_2 | \Lambda^{\otimes 2} | B_1 \rangle\rangle & \langle\langle B_2 | \Lambda^{\otimes 2} | B_2 \rangle\rangle \end{pmatrix} \\
&= \begin{pmatrix} 1 & 0 \\ \langle\langle B_2 | \Lambda^{\otimes 2} | B_1 \rangle\rangle & u(\lambda) \end{pmatrix} \tag{43}
\end{aligned}$$

is a 2×2 matrix. And the two eigenvalues of matrix Φ are 1 and $u(\lambda)$. Thus, there exist constants a and b such that

$$k_A(m) = a + bu(\lambda)^{m-1}. \tag{44}$$

For different m , we estimate the expectation values $k_A(m)$ and fit these values to Eq. (44). Thus, we can robustly and efficiently estimate the unitarity with only a single exponential fitting.

C. Measurable functions via UIRS

In Secs. III A and III B, we have derived the exponential forms of the expectation values as shown in Theorems 1 and 3. We refer to the two different protocols as independent UIRS and identical UIRS. It is worth mentioning that in our protocol, only functions in the form of Eq. (23) can be directly measured with independent UIRS, and only functions in the form of Eq. (E2) can be directly measured with identical UIRS. Due to the existence of the projectors in Eqs. (23) and (E2), the nonlinear function $\text{Tr}(A^T \Lambda^{\otimes 2})$ can only be measured when A is inside the span of projectors. Also, owing to the different classical postprocessing procedures of the independent UIRS protocol and the identical UIRS protocol, the projectors in Eqs. (23) and (E2) differ from each other. There might exist some operators that can only be measured by the independent

UIRS and some other operators that can only be measured by the identical UIRS, which depends on the choice of the gate set \mathbb{G} . Nonetheless, if we assume the matrix exponential fitting is feasible and the random gates are taken from the n -qubit Clifford group, we can show that any operators that identical UIRS can measure can also be measured via independent UIRS. Below, we delve into this discussion in detail.

From Eq. (23), when fixing two irreducible representations τ and τ' and a matrix A , one can obtain the value of $\text{Tr}[(P_j \otimes P_j)A^T(P_i \otimes P_i)\Lambda^{\otimes 2}]$, where P_i and P_j are projectors onto the irreducible representation of τ inside the Liouville representation ω , and $P_{i'}$ and $P_{j'}$ are projectors onto the irreducible representation of τ' inside ω . Thus, independent UIRS can measure observables in the span of $\{P_\tau \otimes P_{\tau'}\}$, where P_τ and $P_{\tau'}$ are projectors onto the irreducible representation space associated with τ and τ' in ω , respectively; τ and τ' are arbitrary irreducible representations of group \mathbb{G} . If ω contains several copies of τ , then $P_\tau = \sum_i P_i$ is the sum of the projectors onto each copy of τ , where P_i is the projector onto the i th copy of τ inside the representation ω . We denote the span of the projectors in this case as $\mathcal{S}_1^{\mathbb{G}} = \text{span}\{P_\tau \otimes P_{\tau'}\}$.

Similarly, from Eq. (E2), identical UIRS can only measure observables in the span of $\{P_\nu\}$, where P_ν is a projector onto the irreducible representation space associated with ν in $\omega^{\otimes 2}$; ν is an irreducible representations of group \mathbb{G} . We denote the span of the projectors in this case as $\mathcal{S}_2^{\mathbb{G}} = \text{span}\{P_\nu\}$.

Note that $\mathcal{S}_1^{\mathbb{G}}$ and $\mathcal{S}_2^{\mathbb{G}}$ are related to the group \mathbb{G} and would be larger if \mathbb{G} is larger. Below, we consider the case that \mathbb{G} is the n -qubit Clifford group. In this case,

$$\mathcal{S}_1^{\mathbb{G}} = \text{span}\{P_{\text{tr}} \otimes P_{\text{tr}}, P_{\text{tr}} \otimes P_{\text{ad}}, P_{\text{ad}} \otimes P_{\text{tr}}, P_{\text{ad}} \otimes P_{\text{ad}}\}; \tag{45}$$

$$\mathcal{S}_2^{\mathbb{G}} = \text{span}\{P_d, P_{id}, P_r, P_l, P_{[S]}, P_{\{S\}}, P_{[A]}, P_{\{A\}}\}. \tag{46}$$

Here, P_{tr} and P_{ad} are two distinct irreducible representations of the Clifford group in ω as introduced in Sec. II B; $P_d, P_{id}, P_r, P_l, P_{[S]}, P_{\{S\}}, P_{[A]},$ and $P_{\{A\}}$ are eight different irreducible representations of the Clifford group in $\omega^{\otimes 2}$ [31]. Then, the operators that can be measured via the independent UIRS and the identical UIRS are determined by Eqs. (45) and (46), respectively.

For OTOC, the first example in this work, the observable A is $(d^2 - 1)^2 \sum_{\sigma \in \mathbb{P}_n^0} \text{Tr}(W\sigma W\sigma) |\sigma \otimes \sigma\rangle\langle V^\dagger \otimes V|$ satisfying $A = P_{\text{ad}}^{\otimes 2} A P_{\text{ad}}^{\otimes 2}$. Meanwhile, $A = P_d A P_d$. Thus, OTOC can be measured by both the independent UIRS and the identical UIRS. But the difference is that the former only requires the single-exponential fitting, and the latter requires the matrix-exponential fitting as P_d contains multiple copies of an irreducible representation of \mathbb{G} . For unitarity, as discussed in the previous section, this quantity can be measured via the identical UIRS. Meanwhile, the unitarity is equal to $\langle\langle B_2 | \Lambda^{\otimes 2} | B_2 \rangle\rangle$, with B_2 defined in Eq. (42), and corresponds to the observable of $|B_2\rangle\langle B_2|$. Note that $|B_2\rangle\langle B_2|$ is inside $P_{\text{ad}} \otimes P_{\text{ad}}$ so unitarity can be also measured via the independent UIRS. The difference between the two protocols is the classical postprocessing and the sample complexity.

Though the two examples in this work can be measured via both protocols, there exist examples that can only be measured via independent UIRS. For instance, the observable $P_{\text{ad}}^{\otimes 2}$ can only be measured via the independent UIRS, as it does not

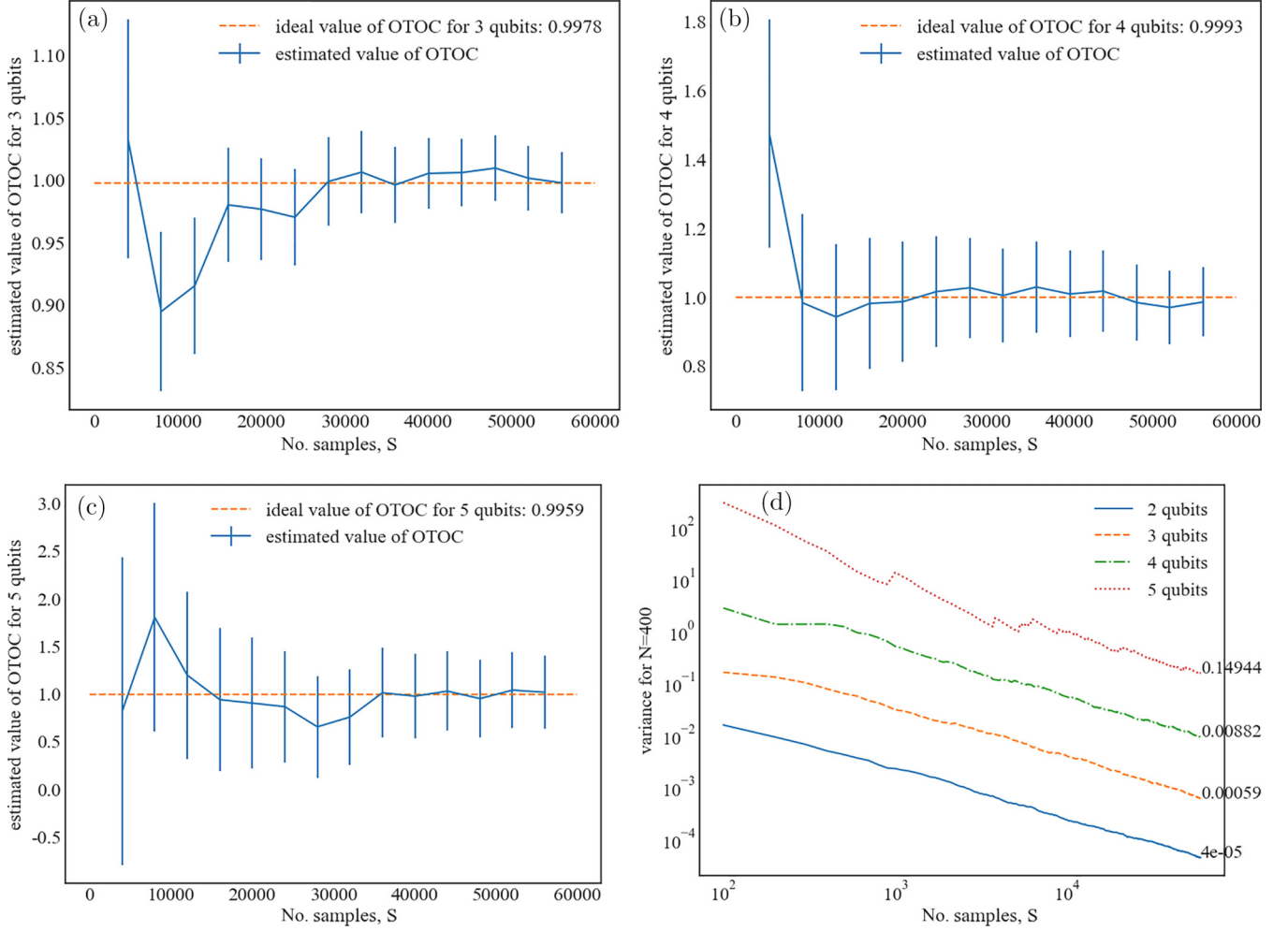


FIG. 3. Convergence of estimated OTOC for (a) three qubits, (b) four qubits, and (c) five qubits. In these figures, the horizontal axis corresponds to S , the number of sampling sequences in $\hat{k}_A(m, S)$. The orange horizontal line denotes the theoretical OTOC value in the ideal case that we want to obtain. And the blue line represents the average \bar{x} over $\{x_1, \dots, x_N\}$, for $N = 400$ and $S_{\max} = 60\,000$ here. Moreover, the blue vertical line represents the standard deviation for $\bar{x}(S)$. Due to the substantial size of the Clifford group, especially for qubit numbers beyond two, exhaustive sampling from the entire Clifford group proves challenging. Instead, we opt to randomly select a subgroup from the n -qubit Clifford group and sample from this subgroup for simulation purposes. Moreover, owing to the extensive size of the Clifford group, the estimate $x_i(S)$ exhibits considerable variation, and $\bar{x}(S)$ deviates from the ideal OTOC when the number of samples S is relatively small. As the sample size S increases, the estimated result gradually converges towards the ideal value. (d) Variance for two, three, four, and five qubits. In this figure, we illustrate the variance of $\bar{x}(S)$. As $\bar{x}(S)$ is the mean value of $\{x_1(S), \dots, x_N(S)\}$, the relationship between their variance is that $\text{Var}[\bar{x}(S)] = \frac{\text{Var}[x_i(S)]}{N}$, where $N = 400$ is the size of the estimates. In this figure, the horizontal axis is the logarithm of the sample number S , and the vertical axis represents the logarithm of the variance for different qubits.

belong to $\mathcal{S}_2^{\mathbb{G}}$. Conversely, we found that there is no example that can only be measured via identical UIRS for the n -qubit Clifford group as $\mathcal{S}_2^{\mathbb{G}} \subseteq \mathcal{S}_1^{\mathbb{G}}$. This can also be easily seen from the following equations:

$$\begin{aligned}
 P_{id} &= P_{\text{tr}} \otimes P_{\text{tr}}, \\
 P_l &= P_{\text{ad}} \otimes P_{\text{tr}}, \\
 P_r &= P_{\text{tr}} \otimes P_{\text{ad}}, \\
 P_d &= P_d \cdot (P_{\text{ad}} \otimes P_{\text{ad}}), \\
 P_{[S]} &= P_{[S]} \cdot (P_{\text{ad}} \otimes P_{\text{ad}}), \\
 P_{\{S\}} &= P_{\{S\}} \cdot (P_{\text{ad}} \otimes P_{\text{ad}}),
 \end{aligned}$$

$$\begin{aligned}
 P_{[A]} &= P_{[A]} \cdot (P_{\text{ad}} \otimes P_{\text{ad}}), \\
 P_{\{A\}} &= P_{\{A\}} \cdot (P_{\text{ad}} \otimes P_{\text{ad}}).
 \end{aligned} \tag{47}$$

Thus, for the n -qubit Clifford group, $\mathcal{S}_2^{\mathbb{G}} \subset \mathcal{S}_1^{\mathbb{G}}$ and we conclude that the observables that independent UIRS can measure contain the observables that identical UIRS can measure. This property comes from the fact the Liouville representation only contains a trivial representation and another irreducible representation for the Clifford group. For the same reason, the property that independent UIRS can measure more observables than identical UIRS holds for any unitary 2-design group [32]. Nonetheless, this phenomenon does not hold for any group. For the Pauli group, identical UIRS can

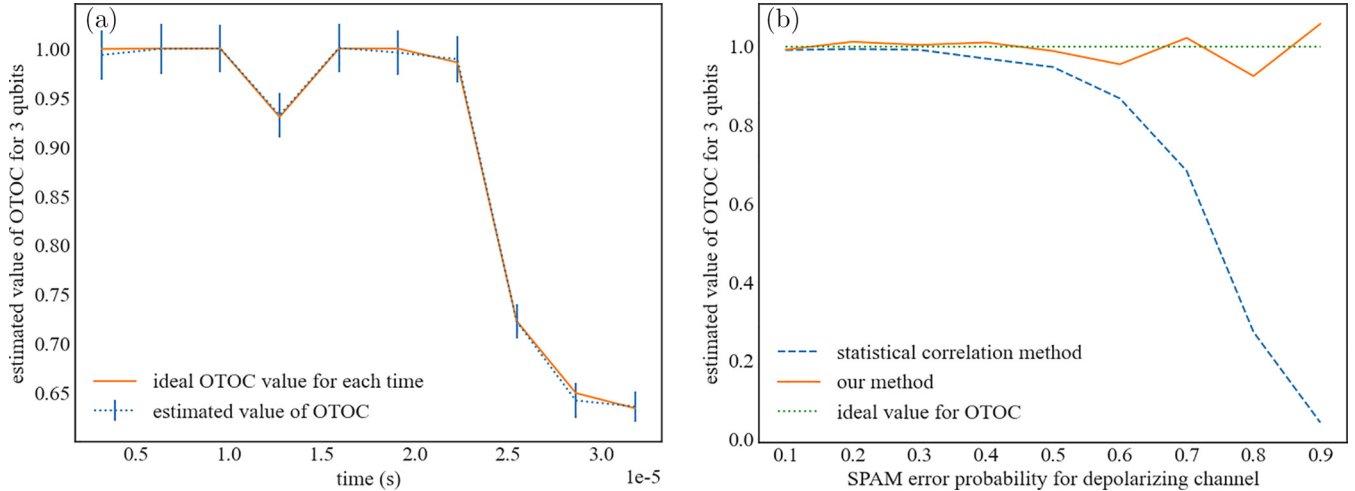


FIG. 4. (a) Variation of OTOC with time. In this figure, the horizontal axis is the time and the orange line is the ideal OTOC value. The blue line is the mean value $\bar{x}(S)$ for $N = 400$, $S = 30\,000$, and the standard deviation of $\bar{x}(S)$ is shown as the error bar. (b) Compared OTOC value under SPAM errors. In this figure, the horizontal axis is the probability p in the depolarizing channel, after which the state ρ undergoes the transformation $(1-p)\rho + \frac{p}{d}\mathbf{1}$; the green line is the ideal OTOC value; the orange line is the estimated value of OTOC from our method; and the blue line is the estimated OTOC from the statistical correlation method [21].

measure more than independent UIRS, as $\mathcal{S}_1^G \subset \mathcal{S}_2^G$ in this case.

IV. SIMULATION RESULTS AND ANALYSIS

In this part, we present the simulation results leveraging the generalized UIRS protocol, primarily focusing on the estimation of OTOC. OTOC is defined with a Hamiltonian evolution and two observables V and W . In our simulation, the evolution is a unitary dynamics, $U_t = e^{-iHt}$, where H is a disordered Ising interaction [33],

$$H_{\text{disordered Ising}} = \sum_{i < j} J_{i,j} \sigma_i^x \sigma_j^x + \frac{B}{2} \sum_i \sigma_i^z + \sum_i \frac{D_i}{2} \sigma_i^z, \quad (48)$$

where $J_{i,j} = \frac{J_0}{|i-j|^\alpha}$ and D_i is uniformly and randomly chosen from $[-D_{\max}, D_{\max}]$. And we set the random Clifford gates to be noiseless in the simulation. The observables V and W are set as Pauli operators with σ_y on the last qubit and σ_x on the next-to-last qubit, respectively. Moreover, the POVM $\{E_x\}$ is chosen as the computational-basis measurement and the initial state ρ is assigned as $|0\rangle\langle 0|$.

As introduced in Eq. (28), to evaluate OTOC, we first calculate the correlation function $f_A(x, y, \mathbf{g}^1, \mathbf{g}^2, m)$ by sampling two sequences of Clifford gates with length m , denoted as $(\mathbf{g}^1, \mathbf{g}^2)$. To reduce the computational difficulties, we simplify the expression of Eq. (28), with the results shown in Appendix F. Then we simulate Eq. (19) by summing over all measurement results x and y to obtain an estimator $\hat{k}_A(m, \mathbf{g}^1, \mathbf{g}^2) = \sum_{x,y} f_A(x, y, \mathbf{g}^1, \mathbf{g}^2, m) p(x, \mathbf{g}^1) p(y, \mathbf{g}^2)$. Then, we repeat sampling $(\mathbf{g}^1, \mathbf{g}^2)$ for S times to obtain the expectation, which we denote as $\hat{k}_A(m, S)$.

As we propose, the value of OTOC is related to the decay of $k_A(m)$. Here, we take m to be 1 and 2 and employ the ratio of $\frac{\hat{k}_A(2)}{\hat{k}_A(1)}$ to derive the OTOC estimate. From the simulation

process above, we obtain one OTOC estimate, $x_1(S) = \frac{\hat{k}_A(2,S)}{\hat{k}_A(1,S)}$. Subsequently, through the repetition of this process, we accumulate a series of estimates, $\{x_1(S), \dots, x_N(S)\}$. The ultimate OTOC estimate is then defined as $\bar{x}(S) = \frac{1}{N} \sum_{i=1}^N x_i(S)$. We illustrate the estimated OTOC for three-, four-, and five-qubit systems in Figs. 3(a)–3(c), respectively. The results show that our protocol can accurately estimate the OTOC. Moreover, if we denote the variance of $\bar{x}(S)$ as s^2 and the variance of $x_i(S)$ as σ^2 , then we have $s^2 = \frac{\sigma^2}{N}$. We show the variance of the estimate $\bar{x}(S)$ with respect to the sampling sequence number S for different qubit systems in Fig. 3(d).

Furthermore, since OTOC depends on the evolution time, we demonstrate how the estimated OTOC changes with time. We sample 10 timestamps with equal intervals and obtain the estimated OTOC for each time, as shown in Fig. 4(a). The results underscore that the estimate captures the temporal evolution of the OTOC, thereby presenting a valuable tool for investigating quantum scrambling phenomena within quantum many-body systems.

As we proposed, our method is robust to SPAM error and we compare it with another method using statistical correlation in [21], as illustrated in Fig. 4(b). The statistical correlation method consists of applying a global unitary to an arbitrary state and then separately measuring $\langle W(t) \rangle$ and $\langle V^\dagger W(t) V \rangle$ to obtain the OTOC. For the SPAM error, we introduce the depolarizing channel after state preparation and before measurement. For the density matrix ρ and the error probability p , the state undergoes a transformation to $(1-p)\rho + \frac{p}{d}\mathbf{1}$ after the depolarizing channel. The simulation result highlights the advantageous performance of our protocol, particularly when the SPAM errors are large.

V. CONCLUSION

In this work, we propose two generalized UIRS protocols that can be used to estimate nonlinear channel properties

robust to SPAM errors. In these protocols, we collect shadow data by applying random gate sequences and POVM measurement and calculate the expectation of the correlation function with shadow data from independent or identical gate sequences. Then, the expectation exhibits an exponential decay against the circuit depth, from which we can extract the nonlinear property of quantum channels. We show the application of our protocols in evaluating OTOC, an important quantity in quantum many-body systems. From the simulation results, we demonstrate the efficiency of our protocol and robustness against SPAM errors. As OTOC is directly related to non-stabilizerness [28], our protocol can be used to measure this essential resource in universal quantum computing. Besides OTOC, we also demonstrate the application of our protocol in estimating other properties such as unitarity, which characterizes the incoherence of quantum channels.

Note that in this work, we only study second-order correlation functions. The whole protocol can be generalized to the case using three or more pieces of shadow data and be used to investigate higher-order channel properties. Meanwhile, we can simultaneously estimate many nonlinear properties from the same shadow data by employing the median of means estimators [6,16]. The sample complexity would have a logarithmic relationship with the number of estimation properties, which we will detail in Appendix G. Moreover, the examples in our work mainly focus on the n -qubit Clifford group. One can investigate the properties derived with groups \mathbb{G} other than the n -qubit Clifford group and explore the possibility that \mathbb{G} is not a group.

ACKNOWLEDGMENTS

The authors would like to thank Wenjun Yu and Pei Zeng for the helpful discussions. This work was supported by the

National Natural Science Foundation of China Grant No. 12174216 and the Innovation Program for Quantum Science and Technology Grants No. 2021ZD0300804 and No. 2021ZD0300702.

APPENDIX A: PROOF OF THEOREM 1

Theorem 1. (Exponential decay). Given a generalized UIRS protocol with group \mathbb{G} , the expectation value $k_A(m)$ of the second-order correlation function given by Eq. (21) has the following form:

$$k_A(m) = \text{Tr}\{\Theta(\{E_x\}, \rho)[\Phi(A, \Lambda)]^{m-1}\}, \quad (\text{A1})$$

where $\Theta(\{E_x\}, \rho)$ and $\Phi(A, \Lambda)$ are induced matrices dependent on SPAM and channel Λ , respectively. If $\omega(g) = \tau_1(g)^{n_{\tau_1}} \oplus \omega'_1(g)$, where $\omega'_1(g)$ contains no copy of $\tau_1(g)$, and $\omega(g) = \tau_2(g)^{n_{\tau_2}} \oplus \omega'_2(g)$, where $\omega'_2(g)$ contains no copy of $\tau_2(g)$, then the matrix $\Phi(A, \Lambda)$ has the element

$$[\Phi(A, \Lambda)]_{i\tilde{i}, j\tilde{j}} = \frac{1}{|P_j||P_{j'}|} \text{Tr}[(P_j \otimes P_{j'})A^T(P_i \otimes P_{i'})\Lambda^{\otimes 2}], \quad (\text{A2})$$

where P_i and P_j are the projectors onto the i th and j th copies of τ_1 inside the representation ω , respectively, and $P_{i'}$ and $P_{j'}$ are the projectors onto the i' th and j' th copies of τ_2 inside the representation ω , respectively. In the special case that the representation ω only has one copy of τ_1 and one copy of τ_2 , the matrix $\Phi(A, \Lambda)$ reduces to a number.

Proof.

$$\begin{aligned} k_A(m) &= \sum_{x,y \in \mathcal{X}} \mathbb{E}_{\mathbf{g}^1, \mathbf{g}^2 \in \mathbb{G}^{xm}} \text{Tr} \left\{ B_{xy} [\tau_1(g_m^1) \otimes \tau_2(g_m^2)] \prod_{i=1}^{m-1} A [\tau_1(g_i^1) \otimes \tau_2(g_i^2)] \right\} \langle \langle \tilde{E}_x | \prod_{i=1}^m \Lambda_L \omega(g_i^1) \Lambda_R | \tilde{\rho} \rangle \rangle \langle \langle \tilde{E}_y | \prod_{i=1}^m \Lambda_L \omega(g_i^2) \Lambda_R | \tilde{\rho} \rangle \rangle \\ &= \sum_{x,y \in \mathcal{X}} \mathbb{E}_{\mathbf{g}^1, \mathbf{g}^2 \in \mathbb{G}^{xm}} \text{Tr} \left\{ [B_{xy} \otimes |\Lambda_R(\tilde{\rho})\rangle\rangle \langle\langle \Lambda_L^*(\tilde{E}_x) | \otimes |\Lambda_R(\tilde{\rho})\rangle\rangle \langle\langle \Lambda_L^*(\tilde{E}_y) |] \tau_1(g_m^1) \otimes \tau_2(g_m^2) \otimes \omega(g_m^1) \otimes \omega(g_m^2) \right. \\ &\quad \left. \times \prod_{i=1}^{m-1} (A \otimes \Lambda^{\otimes 2}) \tau_1(g_i^1) \otimes \tau_2(g_i^2) \otimes \omega(g_i^1) \otimes \omega(g_i^2) \right\} \\ &= \sum_{x,y \in \mathcal{X}} \text{Tr} \{ [B_{xy} \otimes |\Lambda_R(\tilde{\rho})\rangle\rangle \langle\langle \Lambda_L^*(\tilde{E}_x) | \otimes |\Lambda_R(\tilde{\rho})\rangle\rangle \langle\langle \Lambda_L^*(\tilde{E}_y) |] P_{\tau_1 \otimes \tau_2} [P_{\tau_1 \otimes \tau_2} (A \otimes \Lambda^{\otimes 2}) P_{\tau_1 \otimes \tau_2}]^{m-1} \}, \end{aligned} \quad (\text{A3})$$

where $P_{\tau_1 \otimes \tau_2} = P_{\tau_1} \otimes P_{\tau_2} = \mathbb{E}_{g_1 \in \mathbb{G}} [\tau_1(g_1) \otimes \omega(g_1)] \otimes \mathbb{E}_{g_2 \in \mathbb{G}} [\tau_2(g_2) \otimes \omega(g_2)]$ is the corresponding representation average projector. If $\omega(g) = \tau_1(g)^{n_{\tau_1}} \oplus \omega'_1(g)$ and $\omega(g) = \tau_2(g)^{n_{\tau_2}} \oplus \omega'_2(g)$, the rank of $P_{\tau_1 \otimes \tau_2}$ is $n_{\tau_1} n_{\tau_2}$ and $P_{\tau_1 \otimes \tau_2} (A \otimes \Lambda^{\otimes 2}) P_{\tau_1 \otimes \tau_2}$ can be viewed as a matrix Φ with dimension $n_{\tau_1} n_{\tau_2} \times n_{\tau_1} n_{\tau_2}$. The element of Φ is given by

$$\Phi_{i\tilde{i}, j\tilde{j}} = \frac{1}{|P_j||P_{j'}|} \text{Tr}[(P_j \otimes P_{j'})A^T(P_i \otimes P_{i'})\Lambda^{\otimes 2}], \quad (\text{A4})$$

where P_i and P_j are the projectors onto the i th and j th copies of τ_1 inside the representation ω ; $P_{i'}$ and $P_{j'}$ are the projectors onto the i' th and j' th copies of τ_2 inside the representation ω . In particular, if $n_{\tau_1} = n_{\tau_2} = 1$, then the matrix $\Phi(A, \Lambda)$ degenerates into a real number, $\text{Tr}(P_{\tau_1 \otimes \tau_2} A^T P_{\tau_1 \otimes \tau_2} \Lambda^{\otimes 2}) / (|P_{\tau_1}||P_{\tau_2}|)$, with P_{τ_1} and P_{τ_2} being the projectors onto the image of τ_1 and τ_2 inside ω , respectively. ■

APPENDIX B: PROOF OF COROLLARY 1

Corollary 1. The expectation value $k_A(m)$ of the second-order correlation function defined in Eq. (28) has an exponential decay behavior with respect to the sequence length m , $k_A(m) = a[dO(t)]^{m-1}$. The decay parameter is $p(A) = dO(t)$ with $O(t) = \frac{1}{d}\text{Tr}[W^\dagger V(t)^\dagger W V(t)]$ being OTOC.

Proof. From Theorem 1, the function $k_A(m)$ has an exponential decay with decay parameter as $p(A) = \frac{\text{Tr}(P_{\text{ad}}^{\otimes 2} A^T P_{\text{ad}}^{\otimes 2} \mathcal{U}_t^{\otimes 2})}{|P_{\text{ad}}|^2}$ and P_{ad} is the corresponding projector onto the representation τ_{ad} inside ω . Hence, we only need to prove that $p(A) = dO(t)$.

The projector $P_{\text{ad}} = \sum_{\sigma \in \mathbb{P}_n^0} |\sigma\rangle\langle\sigma|$ and $|P_{\text{ad}}| = d^2 - 1$. $P_{\text{ad}}^{\otimes 2} = \sum_{\sigma_1, \sigma_2 \in \mathbb{P}_n^0} |\sigma_1 \otimes \sigma_2\rangle\langle\sigma_1 \otimes \sigma_2|$. For two Pauli operators, there is $\text{Tr}(\sigma_1 \sigma_2) = \delta_{1,2}$. Note that $A^T = (d^2 - 1)^2 \sum_{\sigma \in \mathbb{P}_n^0} \text{Tr}(W \sigma W \sigma) |V \otimes V\rangle\langle\sigma \otimes \sigma|$, where W and V are both nonidentity Pauli operators. Thus, operator A^T is supported on the projection space of $P_{\text{ad}}^{\otimes 2}$,

$$P_{\text{ad}}^{\otimes 2} A^T P_{\text{ad}}^{\otimes 2} = A^T. \quad (\text{B1})$$

The decay parameter now becomes $p(A) = \frac{1}{(d^2-1)^2} \text{Tr}(A^T \mathcal{U}_t^{\otimes 2})$. Before calculating $p(A)$, we first prove that in the expansion, $|F(W \otimes W)\rangle = \sum_{\sigma \in \mathbb{P}_n} \text{Tr}(W \sigma W \sigma) |\sigma \otimes \sigma\rangle$, with F being the SWAP operator. In fact,

$$\begin{aligned} \langle\langle \sigma_1 \otimes \sigma_2 | F(W \otimes W) \rangle\rangle &= \text{Tr}[(\sigma_1 \otimes \sigma_2) F(W \otimes W)] \\ &= \text{Tr}[F(W \sigma_1 \otimes W \sigma_2)] \\ &= \text{Tr}(W \sigma_1 W \sigma_2) \\ &= \text{Tr}(W \sigma_1 W \sigma_1) \delta_{1,2}. \end{aligned} \quad (\text{B2})$$

Since F and W are Hermitian matrices, the operator A^T can be written as

$$\begin{aligned} A^T &= (d^2 - 1)^2 [|V \otimes V\rangle\langle\langle F(W \otimes W) | \\ &\quad - |V \otimes V\rangle\langle\langle \mathbf{1}/\sqrt{d} \otimes \mathbf{1}/\sqrt{d} |]. \end{aligned} \quad (\text{B3})$$

Therefore,

$$\begin{aligned} \frac{\text{Tr}(A^T \mathcal{U}_t^{\otimes 2})}{(d^2 - 1)^2} &= \text{Tr}[|V \otimes V\rangle\langle\langle F(W \otimes W) | \mathcal{U}_t^{\otimes 2}] \\ &\quad - \text{Tr}[|V \otimes V\rangle\langle\langle \mathbf{1} \otimes \mathbf{1} | \mathcal{U}_t^{\otimes 2}] / d \\ &= \langle\langle F(W \otimes W) | \mathcal{U}_t(V) \otimes \mathcal{U}_t(V) \rangle\rangle \\ &\quad - \langle\langle \mathbf{1} \otimes \mathbf{1} | \mathcal{U}_t(V) \otimes \mathcal{U}_t(V) \rangle\rangle / d \\ &= \text{Tr}\{F[W \mathcal{U}_t(V) \otimes W \mathcal{U}_t(V)]\} - |\text{Tr}V|^2 / d \\ &= \text{Tr}[W \mathcal{U}_t(V) W \mathcal{U}_t(V)] = d \cdot O(t). \end{aligned} \quad (\text{B4})$$

Then, $k_A(t)$ has exponential decay with the decay parameter as $p(A) = dO(t)$. \blacksquare

Notice that the above derivation can be straightforwardly generalized to the case that V and W are arbitrary operators by considering the correction term $|\text{Tr}V|^2/d$. Thus, the OTOC estimation protocol can be applied to generic observables V and W .

APPENDIX C: GENERALIZED UIRS WITH APPROXIMATE 2-DESIGN

As elaborated in the previous part, the OTOC can be estimated via the independent UIRS protocol with the Clifford group. In practice, one can use the ϵ -approximate 2-design gate set to replace the Clifford group. Since normally the approximate 2-design gate set is easier to implement, this property can greatly enhance our OTOC estimation practicality. Below, we show the faithfulness of replacing the Clifford group with the approximate 2-design gate set. Specifically, we show that the second-order correlation function $k_A(m)$ for the Clifford group and that for the approximate 2-design group are close.

Recall that the correlation function $k_A(m)$ in the independent UIRS protocol has the following expression:

$$\begin{aligned} k_A(m) &= \text{Tr}[B' P_{\tau_1 \otimes \tau_2} (A' P_{\tau_1 \otimes \tau_2})^{m-1}], \\ A' &= A \otimes \Lambda^{\otimes 2}, \\ B' &= \sum_{x, y \in \mathcal{X}} B_{xy} \otimes |\Lambda_R(\tilde{\rho})\rangle\langle\langle \Lambda_L^*(\tilde{E}_x) | \otimes |\Lambda_R(\tilde{\rho})\rangle\langle\langle \Lambda_L^*(\tilde{E}_y) |. \end{aligned} \quad (\text{C1})$$

We rewrite the above equation into the following form:

$$k_A(m) = \text{Tr}[\Pi(B' \otimes A'^{\otimes m-1}) P_{\tau_1 \otimes \tau_2}^{\otimes m}], \quad (\text{C2})$$

where Π is a permutation rotating the i th part to the $(i+1)$ -th part, and the last part to the first part. Recall that $P_{\tau_1 \otimes \tau_2} = P_{\tau_1} \otimes P_{\tau_2}$, and $\forall i \in \{1, 2\}$,

$$P_{\tau_i} = \mathbb{E}_{g_i \in \mathbb{G}} [\tau_i(g_i) \otimes \omega(g_i)]. \quad (\text{C3})$$

For the Clifford group, the irreducible representation τ_i can be taken as τ_{tr} or τ_{ad} . Taking τ_{ad} as an instance, we can always express $\tau_{\text{ad}}(g_i) = P_{\text{ad}} \omega(g_i) P_{\text{ad}}$, with P_{ad} the projector into the irreducible representation space of τ_{ad} . Given an ϵ -approximate design gate set $\{u_i\}$, we can also define $\tau_{\text{ad}}(u_i) = P_{\text{ad}} \omega(u_i) P_{\text{ad}}$. Note that the ϵ -approximate 2-design gate set $\mathcal{S}^\epsilon = \{u_i\}$ have the following property [30]:

$$\begin{aligned} (1 - \epsilon) \mathbb{E}_{g_i \in \mathcal{C}_n} \omega^{\otimes 2}(g_i) &\leq \mathbb{E}_{u_i \in \mathcal{S}^\epsilon} \omega^{\otimes 2}(u_i) \\ &\leq (1 + \epsilon) \mathbb{E}_{g_i \in \mathcal{C}_n} \omega^{\otimes 2}(g_i). \end{aligned} \quad (\text{C4})$$

We can deduce that

$$(1 - \epsilon) P_{\tau_i}^{\mathcal{C}} \leq P_{\tau_i}^{\mathcal{S}^\epsilon} \leq (1 + \epsilon) P_{\tau_i}^{\mathcal{C}}, \quad (\text{C5})$$

where $P_{\tau_i}^{\mathcal{C}}$ is the projector $\mathbb{E}_{g_i \in \mathcal{C}_n} [\tau_i(g_i) \otimes \omega(g_i)]$ averaging over the n -qubit Clifford group \mathcal{C}_n , and $P_{\tau_i}^{\mathcal{S}^\epsilon}$ is that over the approximate 2-design gate set \mathcal{S}^ϵ . As a consequence, we have

$$(1 - \epsilon)^{2m} (P_{\tau_1 \otimes \tau_2}^{\mathcal{C}})^{\otimes m} \leq (P_{\tau_1 \otimes \tau_2}^{\mathcal{S}^\epsilon})^{\otimes m} \leq (1 + \epsilon)^{2m} (P_{\tau_1 \otimes \tau_2}^{\mathcal{C}})^{\otimes m}. \quad (\text{C6})$$

Hence,

$$\begin{aligned} &|k_A^{\mathcal{S}^\epsilon}(m) - k_A^{\mathcal{C}_n}(m)| \\ &= |\text{Tr}\{\Pi(B' \otimes A'^{\otimes m-1}) [(P_{\tau_1 \otimes \tau_2}^{\mathcal{S}^\epsilon})^{\otimes m} - (P_{\tau_1 \otimes \tau_2}^{\mathcal{C}})^{\otimes m}]\}| \\ &\leq [(1 + \epsilon)^{2m} - 1] |\text{Tr}\{\Pi(B' \otimes A'^{\otimes m-1}) (P_{\tau_1 \otimes \tau_2}^{\mathcal{C}})^{\otimes m}\}| \\ &= [(1 + \epsilon)^{2m} - 1] k_A^{\mathcal{C}_n}(m). \end{aligned} \quad (\text{C7})$$

Thus, the relative error between the correlation functions evaluated by two gate sets is $(1 + \epsilon)^{2m} - 1$. Since m is normally a constant, we only need to set ϵ sufficiently small to make this relative error small enough and enable the final fitting procedure. Thus, the approximate 2-design can replace the Clifford group without compromising the faithfulness of the estimation.

APPENDIX D: SAMPLE COMPLEXITY

In this Appendix, we give a bound of the variance in the OTOC measurement protocol, which also characterizes the sampling complexity. Recall that in the measurement protocol, $B_{xy} = |\rho\rangle\langle E_x| \otimes |\rho\rangle\langle E_y|$ and

$$A = (d^2 - 1)^2 \sum_{\sigma \in \mathbb{P}_n^0} \text{Tr}(W\sigma W\sigma) |\sigma \otimes \sigma\rangle\langle V \otimes V| = (d^2 - 1)^2 [F(W \otimes W)] \langle V \otimes V| - |\mathbf{1}/\sqrt{d} \otimes \mathbf{1}/\sqrt{d}\rangle\langle V \otimes V|, \quad (\text{D1})$$

from which we can directly obtain that $P_{\text{ad}}^{\otimes 2} A P_{\text{ad}}^{\otimes 2} = A$. Define the projector

$$P = \mathbb{E}_{g \in \mathbb{G}} \tau_{\text{ad}}(g)^{\otimes 2} \otimes \omega(g) = P_{\text{ad}}^{\otimes 2} \otimes \mathbf{1} \left[\mathbb{E}_{g \in \mathbb{G}} \omega(g)^{\otimes 3} \right] P_{\text{ad}}^{\otimes 2} \otimes \mathbf{1}. \quad (\text{D2})$$

Since the Clifford group forms a 3-design [34], there is

$$P = \sum_{\pi, \pi' \in S_3} Q_{\pi, \pi'} P_{\text{ad}}^{\otimes 2} \otimes \mathbf{1} |\pi\rangle\langle \pi'| P_{\text{ad}}^{\otimes 2} \otimes \mathbf{1}, \quad (\text{D3})$$

where $Q = (Q_{\pi, \pi'})$ is the Weingarten matrix, and S_3 is the permutation group for three copies of the base Hilbert space, i.e.,

$$\pi |i_1, i_2, i_3\rangle = |i_{\pi(1)}, i_{\pi(2)}, i_{\pi(3)}\rangle. \quad (\text{D4})$$

Theorem 4. Viewing f_A defined in Eq. (28) as a random variable with probability distribution from the shadow data, $p(x, \mathbf{g}^1)p(y, \mathbf{g}^2)$, the variance of this random variable is upper bounded by $O(d^{8m-12})$.

Proof.

$$\begin{aligned} \text{Var}[f_A(m)] &= \mathbb{E}[f_A(m)]^2 - [\mathbb{E}f_A(m)]^2 \\ &\leq \sum_{x,y} \mathbb{E}_{\mathbf{g}^1, \mathbf{g}^2 \in \mathbb{G}^{\times m}} \left(\text{Tr}\{B_{xy}[\tau_1(\mathbf{g}_m^1) \otimes \tau_2(\mathbf{g}_m^2)] \prod_{i=1}^{m-1} A[\tau_1(\mathbf{g}_i^1) \otimes \tau_2(\mathbf{g}_i^2)]\} \right)^2 p(x, \mathbf{g}^1)p(y, \mathbf{g}^2) \\ &= \sum_{x,y} \text{Tr}\{[B_{xy}^{\otimes 2} \otimes |\Lambda_R(\tilde{\rho})\rangle\langle \Lambda_L^*(\tilde{E}_x)| \otimes |\Lambda_R(\tilde{\rho})\rangle\langle \Lambda_L^*(\tilde{E}_y)|] P^{\otimes 2} ([A^{\otimes 2} \otimes \Lambda^{\otimes 2}] P^{\otimes 2})^{m-1}\} \\ &= \sum_{x,y} \text{Tr}[\Theta^{xy} (Q^{\otimes 2} \Omega)^{m-1} Q^{\otimes 2}], \end{aligned} \quad (\text{D5})$$

where Θ^{xy} is the corresponding matrix with element

$$\Theta_{\pi, \pi', \mu, \mu'}^{xy} = \langle \pi' | [P_{\text{ad}}^{\otimes 2} |\rho\rangle\langle E_x| \otimes P_{\text{ad}}^{\otimes 2} \otimes |\Lambda_R(\tilde{\rho})\rangle\langle \Lambda_L^*(\tilde{E}_x)|] | \pi \rangle \langle \mu' | [P_{\text{ad}}^{\otimes 2} |\rho\rangle\langle E_y| \otimes P_{\text{ad}}^{\otimes 2} \otimes |\Lambda_R(\tilde{\rho})\rangle\langle \Lambda_L^*(\tilde{E}_y)|] | \mu \rangle, \quad (\text{D6})$$

and Ω is the corresponding matrix with element

$$\Omega_{\pi, \pi', \mu, \mu'} = \langle \pi' \otimes \mu' | A_{14} \otimes A_{25} \otimes \Lambda_3 \otimes \Lambda_6 | \pi \otimes \mu \rangle. \quad (\text{D7})$$

Here, there are six copies of the base Hilbert space with π, π' being permutation operators of copies 123 and μ, μ' being permutation operators of copies 456. In addition, the subscripts represent the copies that the operator acts on. For example, A_{14} is an operator on the first and the fourth copies of the Hilbert space.

Denote $\|A_{m \times n}\| = \max(m, n) \max_{i,j} |A_{ij}|$ and, obviously,

$$\text{Tr} A_{m \times n} \leq \|A_{m \times n}\|. \quad (\text{D8})$$

Suppose B is a $n \times s$ matrix, then

$$\|AB\| = \max(m, s) \max_{i,k} \left| \sum_j A_{ij} B_{jk} \right| \leq n \max(m, s) \max_{i,j} |A_{ij}| \max_{j,k} |B_{jk}| \leq \max(m, n) \max(n, s) \max_{i,j} |A_{ij}| \max_{j,k} |B_{jk}| = \|A\| \|B\|. \quad (\text{D9})$$

Then, the variance can be further bounded by the inequality

$$\begin{aligned} \mathbb{V}_{f_A(m)} &\leq \left\| \sum_{x,y} \Theta^{xy} (Q^{\otimes 2} \Omega)^{m-1} Q^{\otimes 2} \right\| \leq \|\Theta\| \|Q\|^{2m} \|\Omega\|^{m-1} \leq c \left(\max_{i,j} |\Theta_{ij}| \right) \left(\max_{i,j} |Q_{ij}| \right)^{2m} \left(\max_{i,j} |\Omega_{ij}| \right)^{m-1} \\ &= c \left(\frac{d^2 - 2}{d(d^2 - 1)(d^2 - 4)} \right)^{2m} \left(\max_{i,j} |\Theta_{ij}| \right) \left(\max_{i,j} |\Omega_{ij}| \right)^{m-1} = O(d^{-6m}) \left(\max_{i,j} |\Theta_{ij}| \right) \left(\max_{i,j} |\Omega_{ij}| \right)^{m-1}, \end{aligned} \quad (\text{D10})$$

where c is a constant independent of the system dimension d and $\Theta = \sum_{x,y} \Theta_{xy}$. Thus, we only need to calculate the maximal element of matrices Θ and Ω .

Notice that $\Theta = \tilde{\Theta}^{\otimes 2}$, where the matrix $\tilde{\Theta}$ has the element

$$\tilde{\Theta}_{\pi,\pi'} = \sum_x \langle \langle \pi' | [P_{\text{ad}}^{\otimes 2} |\rho\rangle] \langle E_x |^{\otimes 2} P_{\text{ad}}^{\otimes 2} \otimes |\Lambda_R(\tilde{\rho})\rangle \rangle \langle \Lambda_L^*(\tilde{E}_x) | \rangle | \pi \rangle. \quad (\text{D11})$$

The maximal element of matrix $\tilde{\Theta}$ is of the order of d [16]. Thus, $\max_{i,j} |\Theta_{ij}| = O(d^2)$.

To calculate the matrix Ω , we notice that the projector P_{ad} is supported on the space of traceless matrices, that is,

$$P_{\text{ad}} |\mathbf{1}\rangle = \sum_{\sigma \in \mathbb{P}_n^0} |\sigma\rangle \langle \sigma | \mathbf{1}\rangle = 0, \quad \langle \langle \mathbf{1} | P_{\text{ad}} = \sum_{\sigma \in \mathbb{P}_n^0} \langle \langle \mathbf{1} | \sigma \rangle \langle \sigma | = 0. \quad (\text{D12})$$

The construction of A satisfies that $A = P_{\text{ad}}^{\otimes 2} A P_{\text{ad}}^{\otimes 2}$. Thus, there are four projectors P_{ad} acting on copies 1245 within Eq. (D7). If one of the copies 1245 remains unchanged under a permutation, then the corresponding matrix elements of Ω are zero. For example, we take $\pi = (13)$ and then

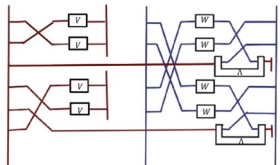
$$P_{\text{ad}}^{\otimes 4} (A_{14} \otimes A_{25}) P_{\text{ad}}^{\otimes 4} \otimes \Lambda_3 \otimes \Lambda_6 | (13) \otimes \mathbf{1}_2 \otimes \mu \rangle = 0. \quad (\text{D13})$$

Therefore, the element $\Omega_{\pi,\pi',\mu,\mu'}$ is nonzero iff $\pi, \pi' \in \{(12), (123), (132)\}$ and $\mu, \mu' \in \{(45), (456), (465)\}$. Next, we leverage the tool of the tensor network to calculate the element of matrix Ω . From Eq. (D1), there is $A^{\otimes 2} = (d^2 - 1)^4 (B + C/d^2 - D/d - E/d)$ with

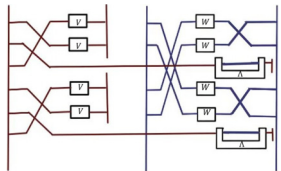
$$B = |F(W \otimes W)^{\otimes 2}\rangle \langle V^{\otimes 4}|, \quad C = |\mathbf{1}^{\otimes 4}\rangle \langle V^{\otimes 4}|, \quad D = |F(W \otimes W) \otimes \mathbf{1}^{\otimes 2}\rangle \langle V^{\otimes 4}|, \quad E = |\mathbf{1}^{\otimes 2} \otimes F(W \otimes W)\rangle \langle V^{\otimes 4}|. \quad (\text{D14})$$

To bound the maximum element of Ω , we need to calculate the maximal element of the matrices generated by B, C , and D , respectively. By symmetry, the elements of D and E have the same order.

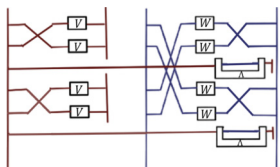
For B , we first take $\pi = (123), \pi' = (12), \mu = \mu' = (456)$ as an example. There is

$$\langle \langle (12) \otimes (456) | B \otimes \Lambda_3 \otimes \Lambda_6 | (123) \otimes (456) \rangle \rangle = \text{Tr} \left(\left(\begin{array}{c} \text{Diagram} \end{array} \right)^2 \right) = d \{ \text{Tr}[\Lambda(\mathbf{1})] \}^2 = d^3. \quad (\text{D15})$$


Similarly,

$$\langle \langle (123) \otimes (456) | B \otimes \Lambda_3 \otimes \Lambda_6 | (12) \otimes (45) \rangle \rangle = \text{Tr} \left(\left(\begin{array}{c} \text{Diagram} \end{array} \right)^2 \right) = d^2 \left(\left(\begin{array}{c} \text{Diagram} \end{array} \right) \right)^2 = d^4. \quad (\text{D16})$$


From the symmetry of the system and these two examples above, we found that the maximum element corresponding to B is obtained when $\pi = \pi' = (12), \mu = \mu' = (45)$. In this case, there is

$$\langle \langle (12) \otimes (45) | B \otimes \Lambda_3 \otimes \Lambda_6 | (12) \otimes (45) \rangle \rangle = \text{Tr}(V^2)^2 \text{Tr}(W^2)^2 \text{Tr}[\Lambda(\mathbf{1})]^2 = d^6. \quad (\text{D17})$$


Similarly, we can obtain, for C ,

$$\max_{\pi,\mu,\pi',\mu'} \langle \langle \pi', \mu' | C \otimes \Lambda_3 \otimes \Lambda_6 | \pi, \mu \rangle \rangle = \langle \langle (12) \otimes (45) | C \otimes \Lambda_3 \otimes \Lambda_6 | (12) \otimes (45) \rangle \rangle = d^6, \quad (\text{D18})$$

and for D ,

$$\max_{\pi, \mu, \pi', \mu'} \langle \langle \pi', \mu' | D \otimes \Lambda_3 \otimes \Lambda_6 | \pi, \mu \rangle \rangle = \langle \langle (12) \otimes (45) | D \otimes \Lambda_3 \otimes \Lambda_6 | (12) \otimes (45) \rangle \rangle = d^5. \quad (\text{D19})$$

Consequently,

$$\mathbf{Var}[f_A(m)] \leq O(d^{8m-12}). \quad (\text{D20})$$

■

From the variance of the single-shot experiment, we first calculate the variance of the estimator given in Eq. (20).
Corollary 3. Suppose the number of samples is S and the estimator for $k_A(m, S)$ is

$$\hat{k}_A(m, S) = \frac{1}{S(S-1)} \sum_{i \neq j} f(x_i, x_j, \mathbf{g}^i, \mathbf{g}^j, m). \quad (\text{D21})$$

Then the variance has an upper bound,

$$\mathbf{Var}[\hat{k}_A(m, S)] \leq \begin{cases} \frac{1}{S(S-1)} O(d^{-4}) + \frac{2(S-2)}{S(S-1)} O(d^{-1}), & m = 1 \\ \frac{1}{S(S-1)} O(d^4) + \frac{2(S-2)}{S(S-1)} O(1), & m = 2 \\ \frac{1}{S(S-1)} O(d^{8m-12}) + \frac{2(S-2)}{S(S-1)} O(d^{4m-4}), & m > 2. \end{cases} \quad (\text{D22})$$

Proof. Denote $s_i = (x_i, \mathbf{g}^i)$ and the variance is

$$\begin{aligned} \mathbf{Var}[\hat{k}_A(m, S)] &= \frac{1}{S^2(S-1)^2} \sum_{i \neq j, i' \neq j'} \mathbf{Cov}[f(s_i, s_j), f(s_{i'}, s_{j'})] \\ &= \frac{1}{S^2(S-1)^2} \left\{ \sum_{i=i', j=j', i \neq j} \mathbf{Cov}[f(s_i, s_j), f(s_{i'}, s_{j'})] + \sum_{i=i', i \neq j, j \neq j', i \neq j'} \mathbf{Cov}[f(s_i, s_j), f(s_{i'}, s_{j'})] \right. \\ &\quad \left. + \sum_{j=j', i \neq j, i \neq i', i' \neq j} \mathbf{Cov}[f(s_i, s_j), f(s_{i'}, s_{j'})] \right\} \\ &= \frac{1}{S^2(S-1)^2} \left\{ \sum_{i \neq j} \mathbf{Var}[f(s_i, s_j)] + 2 \sum_{i \neq j, j \neq j', i \neq j'} \mathbf{Cov}[f(s_i, s_j), f(s_i, s_{j'})] \right\} \\ &= \frac{1}{S(S-1)} \mathbf{Var}[f(s_1, s_2)] + \frac{2(S-2)}{S(S-1)} \mathbf{Cov}[f(s_1, s_2), f(s_1, s_3)] \\ &= \frac{1}{S(S-1)} \mathbf{Var}[f_A(m)] + \frac{2(S-2)}{S(S-1)} \mathbf{Cov}[f(s_1, s_2), f(s_1, s_3)]. \end{aligned} \quad (\text{D23})$$

The first term has already been given, and we only need to calculate the second term. When the sample s_1 is taken as a fixed data, $f(s_1, s_2)$ and $f(s_1, s_3)$ are independent. Therefore,

$$\begin{aligned} \mathbf{Cov}[f(s_1, s_2), f(s_1, s_3)] &= \mathbb{E}_{s_1, s_2, s_3} [f(s_1, s_2) - k_A(m, S)][f(s_1, s_3) - k_A(m, S)] \\ &= \mathbb{E}_{s_1} \left(\mathbb{E}_{s_2, s_3} \{ [f(s_1, s_2) - k_A(m, S)][f(s_1, s_3) - k_A(m, S)] | s_1 \} \right) \\ &= \mathbb{E}_{s_1} \left(\left\{ \mathbb{E}_{s_2} [f(s_1, s_2) | s_1] - k_A(m, S) \right\} \left\{ \mathbb{E}_{s_3} [f(s_1, s_3) | s_1] - k_A(m, S) \right\} \right) \\ &= \mathbf{Var}[f(s_1)], \end{aligned} \quad (\text{D24})$$

where $f(s_1) = \mathbb{E}_{s_2} [f(s_1, s_2) | s_1]$ is the conditional expectation when sample value s_1 is given. Explicitly,

$$\begin{aligned} f(s_1) &= \sum_y \mathbb{E}_{\mathbf{g}^2 \in \mathbb{G}^{\times m}} \text{Tr} \left\{ B_{xy} [\tau_1(\mathbf{g}_m^1) \otimes \tau_2(\mathbf{g}_m^2)] \prod_{i=1}^{m-1} A[\tau_1(\mathbf{g}_i^1) \otimes \tau_2(\mathbf{g}_i^2)] \right\} p(y, \mathbf{g}^2) \\ &= \sum_y \text{Tr} \left\{ [B_{xy} \otimes |\Lambda_R(\tilde{\rho})\rangle\rangle \langle\langle \Lambda_L^*(\tilde{E}_y) |] [\tau_1(\mathbf{g}_m^1) \otimes P'] \prod_{i=1}^{m-1} (\mathbf{1} \otimes P')(A \otimes \Lambda)(\tau_1(\mathbf{g}_i^1) \otimes P') \right\}, \end{aligned} \quad (\text{D25})$$

where

$$\begin{aligned}
P' &= P_{\text{ad}} \otimes \mathbf{1} \left[\mathbb{E}_{g \in \mathbb{G}} \omega(g)^{\otimes 2} \right] P_{\text{ad}} \otimes \mathbf{1} \\
&= P_{\text{ad}} \otimes \mathbf{1} \left(\frac{1}{d^2} |\mathbf{1}^{\otimes 2}\rangle \langle \mathbf{1}^{\otimes 2}| + \frac{1}{d^2 - 1} \sum_{\sigma, \sigma' \in \mathbb{P}_n^0} |\sigma^{\otimes 2}\rangle \langle \sigma'^{\otimes 2}| \right) P_{\text{ad}} \otimes \mathbf{1} \\
&= \frac{1}{d^2 - 1} \sum_{\sigma, \sigma' \in \mathbb{P}_n^0} |\sigma^{\otimes 2}\rangle \langle \sigma'^{\otimes 2}| \\
&= \frac{1}{d^2 - 1} |F - \mathbf{1}^{\otimes 2}/d\rangle \langle F - \mathbf{1}^{\otimes 2}/d| = |B_2\rangle \langle B_2|.
\end{aligned} \tag{D26}$$

Here, we use the fact that the Clifford group forms a 2-design [16].

To calculate the variance of $f(s_1)$, we need to simplify Eq. (D25). Specifically,

$$\begin{aligned}
&(\mathbf{1} \otimes P')(A \otimes \Lambda)(\tau_1(g_i^1) \otimes P') \\
&= \sum_{\alpha, \alpha', \beta, \beta', \sigma \in \mathbb{P}_n^0} \text{Tr}(W\sigma W\sigma)(\mathbf{1} \otimes |\alpha^{\otimes 2}\rangle \langle \alpha'^{\otimes 2}|)(|\sigma \otimes \sigma\rangle \langle V \otimes V | \otimes \Lambda)[\tau_1(g_i^1) \otimes |\beta^{\otimes 2}\rangle \langle \beta'^{\otimes 2}|] \\
&= \sum_{\alpha, \alpha', \beta, \beta', \sigma \in \mathbb{P}_n^0} \text{Tr}(W\sigma W\sigma) \text{Tr}(\sigma \alpha') \text{Tr}(V\beta) \text{Tr}(\alpha' \Lambda(\beta)) [|\sigma\rangle \langle V | \tau_1(g_i^1) \otimes |\alpha\rangle \langle \beta' | \otimes |\alpha\rangle \langle \beta' |] \\
&= \sum_{\alpha, \beta', \sigma \in \mathbb{P}_n^0} \text{Tr}(W\sigma W\sigma) \text{Tr}[\sigma \Lambda(V)] [|\sigma\rangle \langle V | \tau_1(g_i^1) \otimes |\alpha\rangle \langle \beta' | \otimes |\alpha\rangle \langle \beta' |] \\
&= \sum_{\sigma \in \mathbb{P}_n^0} (d^2 - 1) \text{Tr}(W\sigma W\sigma) \text{Tr}[\sigma \Lambda(V)] |\sigma\rangle \langle V | \tau_1(g_i^1) \otimes P' = A' \tau_1(g_i^1) \otimes P',
\end{aligned} \tag{D27}$$

where $A' = (d^2 - 1) \sum_{\sigma \in \mathbb{P}_n^0} \text{Tr}(W\sigma W\sigma) \text{Tr}[\sigma \Lambda(V)] |\sigma\rangle \langle V |$. Therefore,

$$\begin{aligned}
f(s_1) &= \sum_y \text{Tr} \left\{ \left[|\rho\rangle \langle E_x | \tau_1(g_m^1) \prod_{i=1}^{m-1} A' \tau_1(g_i^1) \right] \otimes [|\rho \otimes \Lambda_R(\tilde{\rho})\rangle \langle E_y \otimes \Lambda_L^*(\tilde{E}_y) | P'] \right\}, \\
&= \sum_y \frac{1}{d^2 - 1} \left\{ \text{Tr}[\rho \Lambda_R(\tilde{\rho})] - \frac{1}{d} \text{Tr}[\Lambda_R(\tilde{\rho})] \right\} \left\{ \text{Tr}[E_y \Lambda_L^*(\tilde{E}_y)] - \frac{1}{d} \text{Tr} E_y \text{Tr}[\Lambda_L^*(\tilde{E}_y)] \right\} \text{Tr} \left[|\rho\rangle \langle E_x | \tau_1(g_m^1) \prod_{i=1}^{m-1} A' \tau_1(g_i^1) \right] \\
&= O(d^{-1}) \text{Tr} \left[|\rho\rangle \langle E_x | \tau_1(g_m^1) \prod_{i=1}^{m-1} A' \tau_1(g_i^1) \right].
\end{aligned} \tag{D28}$$

When $m = 1$, there is

$$\begin{aligned}
\text{Var}[f(s_1)] &\leq O(d^{-2}) \sum_x \mathbb{E}_{g \in \mathbb{G}} \langle \langle E_x | \tau(g) | \rho \rangle \rangle^2 \langle \langle \Lambda_L^*(\tilde{E}_x) | \omega(g) | \Lambda_R(\tilde{\rho}) \rangle \rangle \\
&= O(d^{-2}) \sum_x \langle \langle E_x^{\otimes 2} \otimes \Lambda_L^*(\tilde{E}_x) | P | \rho^{\otimes 2} \otimes \Lambda_R(\tilde{\rho}) \rangle \rangle \\
&= O(d^{-2}) \sum_x \sum_{\pi \in S_3} \mathcal{Q}_{\pi, \pi} \langle \langle (E_x P_{\text{ad}})^{\otimes 2} \otimes \Lambda_L^*(\tilde{E}_x) | \pi \rangle \rangle \langle \langle \pi | (P_{\text{ad}} \rho)^{\otimes 2} \otimes \Lambda_R(\tilde{\rho}) \rangle \rangle \\
&= O(d^{-2}) \text{Tr} \tilde{\Theta} = O(d^{-1}),
\end{aligned} \tag{D29}$$

where the matrix $\tilde{\Theta}$ is defined in Eq. (D11) and the elements of this matrix have the order $O(d)$.

When $m \geq 2$, according to the conclusion in Ref. [16], the variance for $f(s_1)$ has an upper bound,

$$\text{Var}[f(s_1)] \leq O(d^{-4}) [11u(A')(r(A')^{m-2} + [2(m-2)^2 r(A')^{m-3}] \max\{11u(A'), [11u(A')]^2\})], \tag{D30}$$

where $u(A') = \text{Tr}(A'A'^\dagger)/(d^2 - 1)$ and $r(A') = u(A')(1 + 16d^{-1/3})$. Since

$$\begin{aligned}\text{Tr}(A'A'^\dagger) &= (d^2 - 1)^2 \sum_{\sigma, \sigma' \in \mathbb{P}_n^0} \text{Tr}(W\sigma W\sigma) \text{Tr}(W\sigma' W\sigma') \text{Tr}[\sigma \Lambda(V)] \text{Tr}[\sigma' \Lambda(V)] \text{Tr}(\sigma\sigma') \text{Tr}(V^2) \\ &= (d^2 - 1)^2 d \sum_{\sigma \in \mathbb{P}_n^0} \{\text{Tr}(W\sigma W\sigma) \text{Tr}[\sigma \Lambda(V)]\}^2 = (d^2 - 1)^2 d \sum_{\sigma \in \mathbb{P}_n^0} \text{Tr}[\sigma^{\otimes 2} \Lambda(V)^{\otimes 2}] \\ &= (d^2 - 1)^2 d \text{Tr}[(F - \mathbf{1}^{\otimes 2}/d) \Lambda(V)^{\otimes 2}] = (d^2 - 1)^2 d \left(\text{Tr}[\Lambda(V)^2] - \frac{1}{d} \{\text{Tr}[\Lambda(V)]\}^2 \right) = O(d^6),\end{aligned}\quad (\text{D31})$$

we can obtain $u(A') = O(d^4)$ and $r(A') = O(d^4)$. When $m > 2$, the variance can be bounded by $O(d^{4m-4})$. When $m = 2$, the variance can be bounded by $O(1)$, which implies that the variance is bounded by a constant. From Eq. (D23), the overall variance is

$$\text{Var}[\hat{k}_A(m, S)] \leq \begin{cases} \frac{1}{S(S-1)} O(d^{-4}) + \frac{2(S-2)}{S(S-1)} O(d^{-1}), & m = 1 \\ \frac{1}{S(S-1)} O(d^4) + \frac{2(S-2)}{S(S-1)} O(1), & m = 2 \\ \frac{1}{S(S-1)} O(d^{8m-12}) + \frac{2(S-2)}{S(S-1)} O(d^{4m-4}), & m > 2. \end{cases}\quad (\text{D32})$$

■

Next, we calculate the variance of OTOC evaluated by the ratio of $k_A(2)$ and $k_A(1)$ as shown below,

$$\hat{O} = \frac{\hat{k}_A(2, S)}{\hat{k}_A(1, S)} = \frac{\frac{1}{S(S-1)} \sum_{i \neq j} f(x_i, x_j, \mathbf{g}^i, \mathbf{g}^j, 2)}{\frac{1}{S(S-1)} \sum_{i \neq j} f(x_i, x_j, \mathbf{g}^i, \mathbf{g}^j, 1)},\quad (\text{D33})$$

where we assume that the numbers of samples for two cases are both equal to S . For an arbitrary m , the expectation value of $\hat{k}_A(m, S)$ is

$$\begin{aligned}\mathbb{E} \hat{k}_A(m, S) &= k_A(m, S) = \sum_{x, y} \mathbb{E}_{\mathbf{g}^1, \mathbf{g}^2 \in \mathbb{G}^{\times m}} \left(\text{Tr} \left\{ B_{xy} [\tau_1(g_m^1) \otimes \tau_2(g_m^2)] \prod_{i=1}^{m-1} A[\tau_1(g_i^1) \otimes \tau_2(g_i^2)] \right\} \right) p(x, \mathbf{g}^1) p(y, \mathbf{g}^2) \\ &= \sum_{x, y} \text{Tr} \{ [B_{xy} \otimes |\Lambda_R(\tilde{\rho})\rangle\rangle \langle\langle \Lambda_L^*(\tilde{E}_x) | \otimes |\Lambda_R(\tilde{\rho})\rangle\rangle \langle\langle \Lambda_L^*(\tilde{E}_y) |] P'^{\otimes 2} ([A \otimes \Lambda^{\otimes 2}] P'^{\otimes 2})^{m-1} \}.\end{aligned}\quad (\text{D34})$$

When $m = 1$, there is

$$\begin{aligned}k_A(1, S) &= \langle\langle B_2 | \rho \otimes \Lambda_R(\tilde{\rho}) \rangle\rangle^2 \left[\sum_x \langle\langle E_x \otimes \Lambda_L^*(\tilde{E}_x) | B_2 \rangle\rangle \right]^2 \\ &= O(d^{-4}) \left\{ \text{Tr}[\rho \Lambda_R(\tilde{\rho})] - \frac{1}{d} \text{Tr}[\Lambda_R(\tilde{\rho})] \right\}^2 \left\{ \sum_x \text{Tr}[E_x \Lambda_L^*(\tilde{E}_x)] - \frac{1}{d} \sum_x \text{Tr} E_x \text{Tr}[\Lambda_L^*(\tilde{E}_x)] \right\}^2 = O(d^{-2}).\end{aligned}\quad (\text{D35})$$

When $m = 2$, there is

$$\begin{aligned}k_A(2, S) &= k_A(1, S) \langle\langle B_2^{\otimes 2} | A \otimes \Lambda^{\otimes 2} | B_2^{\otimes 2} \rangle\rangle \\ &= O(d^{-2}) \left\{ \langle\langle B_2^{\otimes 2} | [F(W \otimes W)] \rangle\rangle \langle\langle V \otimes V | \otimes \Lambda^{\otimes 2} | B_2^{\otimes 2} \rangle\rangle - \frac{1}{d} \langle\langle B_2^{\otimes 2} | [|\mathbf{1} \otimes \mathbf{1}\rangle\rangle \langle\langle V \otimes V | \otimes \Lambda^{\otimes 2} | B_2^{\otimes 2} \rangle\rangle \right\} \\ &= O(d^{-2}) \langle\langle B_2^{\otimes 2} | [F(W \otimes W)] \rangle\rangle \langle\langle V \otimes V | \otimes \Lambda^{\otimes 2} | B_2^{\otimes 2} \rangle\rangle \\ &= O(d^{-2}) \langle\langle \Lambda \rangle\rangle \langle\langle \Lambda \rangle\rangle \\ &= O(d^{-2}) \text{Tr}[W \Lambda(V) W \Lambda(V)] = O(d^{-1}).\end{aligned}\quad (\text{D36})$$

For two random variables X, Y , the uncertainty of the division X/Y is

$$\Delta \left(\frac{X}{Y} \right) \approx \left(\frac{\mu_X}{\mu_Y} \right)^2 \left[\frac{\sigma_X^2}{\mu_X^2} - 2 \frac{\text{Cov}(X, Y)}{\mu_X \mu_Y} + \frac{\sigma_Y^2}{\mu_Y^2} \right],\quad (\text{D37})$$

where μ_X and μ_Y are the expectation values of X and Y , respectively; σ_X^2 and σ_Y^2 are the variance values of X and Y , respectively. In our case, $X = \hat{k}_A(2, S)$ and $Y = \hat{k}_A(1, S)$. According to Eqs. (D22), (D35), and (D36), there is

$$\mu_X^2 = O(d^{-2}), \quad \mu_Y^2 = O(d^{-4}), \quad \sigma_X^2 = O(d^4 S^{-2}) + O(S^{-1}), \quad \sigma_Y^2 = O(d^{-4} S^{-2}) + O(d^{-1} S^{-1}). \quad (\text{D38})$$

Thus, the uncertainty for the estimator can be bounded by

$$\text{Var}\left(\frac{\hat{k}_A(2, S)}{\hat{k}_A(1, S)}\right) = \text{Var}\left(\frac{X}{Y}\right) \leq \left(\frac{\mu_X}{\mu_Y}\right)^2 \left[\frac{\sigma_X^2}{\mu_X^2} + \frac{\sigma_Y^2}{\mu_Y^2}\right] = O\left(\frac{d^8}{S^2}\right) + O\left(\frac{d^5}{S}\right). \quad (\text{D39})$$

APPENDIX E: PROOF OF THEOREM 3

Theorem 3. (Exponential decay). Given a generalized UIRS protocol with respect to the group \mathbb{G} , the expectation value $k_A(m)$ defined with Eqs. (32) and (33) has the following form:

$$k_A(m) = \text{Tr}(\Theta(\{E_x\}, \rho)[\Phi(A, \Lambda)]^{m-1}), \quad (\text{E1})$$

where $\Theta(\{E_x\}, \rho)$ and $\Phi(A, \Lambda)$ are induced matrices dependent on SPAM and channel Λ , respectively. The matrix $\Phi(A, \Lambda) = P_\tau(A \otimes \Lambda^{\otimes 2})P_\tau$ has the element

$$[\Phi(A, \Lambda)]_{i,j} = \frac{1}{|P_j|} \text{Tr}(P_j A^T P_i \Lambda^{\otimes 2}), \quad (\text{E2})$$

where $P_\tau = \mathbb{E}_{g \in \mathbb{G}} \tau(g) \otimes \omega(g)^{\otimes 2}$, and P_i and P_j are the projectors onto the i th and j th copies of τ inside $\omega^{\otimes 2}$, respectively. In the special case that the representation $\omega^{\otimes 2}$ only has one copy of τ , the matrix $\Phi(A, \Lambda)$ reduces into a number.

Proof.

$$\begin{aligned} k_A(m) &= \sum_{x,y \in \mathcal{X}} \mathbb{E}_{\mathbf{g} \in \mathbb{G}^{\times m}} \text{Tr} \left[B_{xy} \tau(g_m) \prod_{i=1}^{m-1} A \tau(g_i) \right] \langle \tilde{E}_x | \prod_{i=1}^m \Lambda_L \omega(g_i) \Lambda_R | \tilde{\rho} \rangle \langle \tilde{E}_y | \prod_{i=1}^m \Lambda_L \omega(g_i) \Lambda_R | \tilde{\rho} \rangle \\ &= \sum_{x,y \in \mathcal{X}} \mathbb{E}_{\mathbf{g} \in \mathbb{G}^{\times m}} \text{Tr} \left([B_{xy} \otimes |\Lambda_R(\tilde{\rho})\rangle] \langle \Lambda_L^*(\tilde{E}_x) | \otimes |\Lambda_R(\tilde{\rho})\rangle \langle \Lambda_L^*(\tilde{E}_y) | \tau(g_m) \otimes \omega(g_m)^{\otimes 2} \prod_{i=1}^{m-1} \{A \otimes \Lambda^{\otimes 2}[\tau(g_i) \otimes \omega(g_i)^{\otimes 2}]\} \right) \\ &= \sum_{x,y \in \mathcal{X}} \text{Tr} \left([B_{xy} \otimes |\Lambda_R(\tilde{\rho})\rangle] \langle \Lambda_L^*(\tilde{E}_x) | \otimes |\Lambda_R(\tilde{\rho})\rangle \langle \Lambda_L^*(\tilde{E}_y) | \left\{ \mathbb{E}_{g \in \mathbb{G}} [\tau(g) \otimes \omega(g)^{\otimes 2}] (A \otimes \Lambda^{\otimes 2}) \mathbb{E}_{g \in \mathbb{G}} [\tau(g) \otimes \omega(g)^{\otimes 2}] \right\}^{m-1} \right) \\ &= \sum_{x,y \in \mathcal{X}} \text{Tr} \{ [B_{xy} \otimes |\Lambda_R(\tilde{\rho})\rangle] \langle \Lambda_L^*(\tilde{E}_x) | \otimes |\Lambda_R(\tilde{\rho})\rangle \langle \Lambda_L^*(\tilde{E}_y) | [P_\tau(A \otimes \Lambda^{\otimes 2})P_\tau]^{m-1} \}, \end{aligned} \quad (\text{E3})$$

where $P_\tau = \mathbb{E}_{g \in \mathbb{G}} \tau(g) \otimes \omega(g)^{\otimes 2}$ is a projector. The decay parameter $P_\tau(A \otimes \Lambda^{\otimes 2})P_\tau$ can be viewed as a matrix Φ with dimension $\text{rank}(P_\tau) \times \text{rank}(P_\tau)$ when restricting in the space associated with projector P_τ . The matrix Φ has the element

$$[\Phi]_{i,j} = \frac{1}{|P_j|} \text{Tr}(P_j A^T P_i \Lambda^{\otimes 2}), \quad (\text{E4})$$

where P_i is the projector onto the i th copy of τ inside $\omega^{\otimes 2}$. ■

APPENDIX F: FORMULA SIMPLIFICATION OF EQ. (28)

In this section, we simplify the computation of Eq. (28):

$$f_A(x, y, \mathbf{g}^1, \mathbf{g}^2, m) = \text{Tr} \left[B_{xy} \tau(g_m^1) \otimes \tau(g_m^2) \prod_{i=1}^{m-1} A \tau(g_i^1) \otimes \tau(g_i^2) \right]. \quad (\text{F1})$$

For the Clifford group element g , with slight abuse of notations, we denote $g = \omega(g) = U_g$ and $\tau_{\text{ad}}(g) = \tau(g)$. Then, $g = \tau(g) + \frac{1}{d} |\mathbf{1}\rangle \langle \mathbf{1}|$. When $m = 1$, we have

$$\begin{aligned} f_A(x, y, \mathbf{g}^1, \mathbf{g}^2, m) &= \text{Tr} [B_{xy} \tau(g_m^1) \otimes \tau(g_m^2)] = \text{Tr} [|\rho\rangle \langle E_x | \otimes |\rho\rangle \langle E_y | \tau(g^1) \otimes \tau(g^2)] \\ &= \text{Tr} [|\rho\rangle \langle E_x | \tau(g^1)] \cdot \text{Tr} [|\rho\rangle \langle E_y | \tau(g^2)] = \langle E_x | \tau(g^1) | \rho \rangle \cdot \langle E_y | \tau(g^2) | \rho \rangle \\ &= \langle E_x | \left[\omega(g^1) - \frac{1}{d} |\mathbf{1}\rangle \langle \mathbf{1}| \right] | \rho \rangle \cdot \langle E_y | \left[\omega(g^2) - \frac{1}{d} |\mathbf{1}\rangle \langle \mathbf{1}| \right] | \rho \rangle \\ &= \left[\text{Tr}(E_x g^1 \rho g^{1\dagger}) - \frac{1}{d} \right] \cdot \left[\text{Tr}(E_y g^2 \rho g^{2\dagger}) - \frac{1}{d} \right]. \end{aligned} \quad (\text{F2})$$

For $m \geq 2$, $A = (d^2 - 1)^2 \sum_{\sigma \in \mathbb{P}_n^0} \text{Tr}(W\sigma W\sigma) |\sigma \otimes \sigma\rangle \langle V \otimes V|$,

$$\begin{aligned}
f_A(x, y, \mathbf{g}^1, \mathbf{g}^2, m) &= \text{Tr} \left[B_{xy} \tau(g_m^1) \otimes \tau(g_m^2) \prod_{i=1}^{m-1} A \tau(g_i^1) \otimes \tau(g_i^2) \right] \\
&= \text{Tr} \left[B_{xy} \tau(g_m^1) \otimes \tau(g_m^2) \prod_{i=1}^{m-1} (d^2 - 1)^2 \sum_{\sigma \in \mathbb{P}_n^0} \text{Tr}(W\sigma W\sigma) |\sigma \otimes \sigma\rangle \langle V \otimes V | \tau(g_i^1) \otimes \tau(g_i^2) \right] \\
&= (d^2 - 1)^{2(m-1)} \cdot \text{Tr} \left[B_{xy} \tau(g_m^1) \otimes \tau(g_m^2) \sum_{\sigma \in \mathbb{P}_n^0} \text{Tr}(W\sigma W\sigma) |\sigma \otimes \sigma\rangle \langle V \otimes V | \tau(g_1^1) \otimes \tau(g_1^2) \right] \\
&\quad \times \left[\prod_{i=2}^{m-1} \langle V \otimes V | \tau(g_i^1) \otimes \tau(g_i^2) \sum_{\sigma \in \mathbb{P}_n^0} \text{Tr}(W\sigma W\sigma) |\sigma \otimes \sigma\rangle \right]. \tag{F3}
\end{aligned}$$

For an arbitrary normalized Pauli operator σ , there is $g(\sigma) = g^\dagger \sigma g$, $g^{-1}(\sigma) = g \sigma g^\dagger$. We have that $\langle \sigma | \tau(g) = \langle g(\sigma) |$, and thus,

$$\begin{aligned}
\tau(g_m^1) \otimes \tau(g_m^2) \sum_{\sigma \in \mathbb{P}_n^0} \text{Tr}(W\sigma W\sigma) |\sigma \otimes \sigma\rangle &= \sum_{\sigma \in \mathbb{P}_n^0} \text{Tr}(W\sigma W\sigma) |g_m^{1-1}(\sigma) \otimes g_m^{2-1}(\sigma)\rangle, \\
\langle V^\dagger \otimes V | \tau(g_i^1) \otimes \tau(g_i^2) &= \langle g_i^1(V) \otimes g_i^2(V) |. \tag{F4}
\end{aligned}$$

Denote $D = (d^2 - 1)^{2(m-1)}$, $\delta_i = \delta_{g_i^1(V), g_i^2(V)}$. Combine Eqs. (F4) and (F3); then we can obtain that

$$\begin{aligned}
f_A(x, y, \mathbf{g}^1, \mathbf{g}^2, m) &= D \text{Tr} \left[B_{xy} \sum_{\sigma \in \mathbb{P}_n^0} \text{Tr}(W\sigma W\sigma) |g_m^{1-1}(\sigma) \otimes g_m^{2-1}(\sigma)\rangle \langle g_1^1(V) \otimes g_1^2(V) | \right] \\
&\quad \times \prod_{i=2}^{m-1} \langle g_i^1(V) \otimes g_i^2(V) | \sum_{\sigma \in \mathbb{P}_n^0} \text{Tr}(W\sigma W\sigma) |\sigma \otimes \sigma\rangle \\
&= D \text{Tr} \left[B_{xy} \sum_{\sigma \in \mathbb{P}_n^0} \text{Tr}(W\sigma W\sigma) |g_m^{1-1}(\sigma) \otimes g_m^{2-1}(\sigma)\rangle \langle g_1^1(V) \otimes g_1^2(V) | \prod_{i=2}^{m-1} \delta_{g_i^1(V), g_i^2(V)} \text{Tr}[W g_i^1(V)]^2 \right] \\
&= D \left\{ \prod_{i=2}^{m-1} \delta_i \text{Tr}[W g_i^1(V)]^2 \right\} \text{Tr} \left[|\rho\rangle \langle E_x| \otimes |\rho\rangle \langle E_y| \sum_{\sigma \in \mathbb{P}_n^0} \text{Tr}(W\sigma W\sigma) |g_m^{1-1}(\sigma) \otimes g_m^{2-1}(\sigma)\rangle \langle g_1^1(V) \otimes g_1^2(V) | \right] \\
&= D \left\{ \prod_{i=2}^{m-1} \delta_i \text{Tr}[W g_i^1(V)]^2 \right\} \langle E_x \otimes E_y | \sum_{\sigma \in \mathbb{P}_n^0} \text{Tr}(W\sigma W\sigma) |g_m^{1-1}(\sigma) \otimes g_m^{2-1}(\sigma)\rangle \langle g_1^1(V) \otimes g_1^2(V) | |\rho \otimes \rho\rangle \\
&= D \left\{ \prod_{i=2}^{m-1} \delta_i \text{Tr}[W g_i^1(V)]^2 \right\} \langle g_m^1(E_x) \otimes g_m^2(E_y) | \sum_{\sigma \in \mathbb{P}_n^0} \text{Tr}(W\sigma W\sigma) |\sigma \otimes \sigma\rangle \langle g_1^1(V) \otimes g_1^2(V) | |\rho \otimes \rho\rangle \\
&= D \left\{ \prod_{i=2}^{m-1} \delta_i \text{Tr}[W g_i^1(V)]^2 \right\} \langle g_m^1(E_x) \otimes g_m^2(E_y) | \left(|S(W^\dagger \otimes W)\rangle \rangle - \frac{1}{d} |I \otimes I\rangle \right) \langle g_1^1(V) \otimes g_1^2(V) | |\rho \otimes \rho\rangle \\
&= D \left\{ \prod_{i=2}^{m-1} \delta_i \text{Tr}[W g_i^1(V)]^2 \right\} \left\{ \text{Tr}[W^\dagger g_m^1(E_x) W g_m^2(E_y)] - \frac{1}{d} \right\} \text{Tr}[g_1^1(V) \rho] \text{Tr}[g_1^2(V) \rho]. \tag{F5}
\end{aligned}$$

With the equation above, one can quickly evaluate the value of $f_A(x, y, \mathbf{g}^1, \mathbf{g}^2, m)$.

APPENDIX G: SIMULTANEOUS ESTIMATION OF MANY OBSERVABLES

Here, we discuss that our methods can be used to simultaneously estimate multiple nonlinear observables. We take the independent UIRS as an example. Concretely, we assume that the to-be-estimated nonlinear properties correspond to a set \mathcal{A} of

operators and, for each $A \in \mathcal{A}$, the second-order correlation function is

$$f_A(x, y, \mathbf{g}^1, \mathbf{g}^2, m) = \text{Tr}\{B_{xy} \left[\tau_1(g_m^1) \otimes \tau_2(g_m^2) \prod_{i=1}^{m-1} A[\tau_1(g_i^1) \otimes \tau_2(g_i^2)] \right]\}. \quad (\text{G1})$$

Assuming that we obtain $S = NK$ samples in experiments, we can divide data $\{x_i\}_{i=1}^S$ into K groups, and each group contains N samples. For each observable f_A , the median-of-means estimator is defined as

$$\hat{f}_A(N, K) = \text{Median}\{\hat{f}_A^{(1)}(N, 1), \hat{f}_A^{(2)}(N, 1), \dots, \hat{f}_A^{(K)}(N, 1)\}, \quad \text{where} \quad (\text{G2})$$

$$\hat{f}_A^{(k)}(N, 1) = \frac{1}{N(N-1)} \sum_{i \neq j, i, j \in \{N(k-1)+1, \dots, Nk\}} f_A(x_i, x_j, \mathbf{g}^i, \mathbf{g}^j, m), \quad \forall 1 \leq k \leq K. \quad (\text{G3})$$

Note that [6]

$$\text{Var}[\hat{f}_A^{(k)}(N, 1)] \leq \frac{2}{N} \left\{ \text{Var}[f_A^1(y, \mathbf{g}^2, m)] + \text{Var}[f_A^2(x, \mathbf{g}^1, m)] + \frac{1}{N} \text{Var}[f_A(x, y, \mathbf{g}^1, \mathbf{g}^2, m)] \right\}, \quad (\text{G4})$$

where

$$f_A^1(y, \mathbf{g}^2, m) = \mathbb{E}_{x, \mathbf{g}^1} f_A(x, y, \mathbf{g}^1, \mathbf{g}^2, m), \quad (\text{G5})$$

$$f_A^2(x, \mathbf{g}^1, m) = \mathbb{E}_{y, \mathbf{g}^2} f_A(x, y, \mathbf{g}^1, \mathbf{g}^2, m). \quad (\text{G6})$$

Utilizing the property of the median-of-means estimator, for the level of confidence δ and error ϵ , it can be shown that [6] when

$$K = \lceil 2 \ln(2|\mathcal{A}|/\delta) \rceil, \quad (\text{G7})$$

$$N = \left\lceil 34/\epsilon^2 \times 8 \times \max_{A \in \mathcal{A}} \max \left\{ \text{Var}[f_A^1(y, \mathbf{g}^2, m)], \text{Var}[f_A^2(x, \mathbf{g}^1, m)], \sqrt{\text{Var}[f_A(x, y, \mathbf{g}^1, \mathbf{g}^2, m)]} \right\} \right\rceil,$$

there is

$$\Pr \left[\max_{A \in \mathcal{A}} |\hat{f}_A - \mathbb{E}(f_A)| \leq \epsilon \right] = 1 - \delta. \quad (\text{G8})$$

This indicates that we can simultaneously estimate M properties of the channel while only $O(\ln M)$ samples are needed.

-
- [1] B. Swingle, Unscrambling the physics of out-of-time-order correlators, *Nat. Phys.* **14**, 988 (2018).
- [2] J. Wallman, C. Granade, R. Harper, and S. T. Flammia, Estimating the coherence of noise, *New J. Phys.* **17**, 113020 (2015).
- [3] V. P. Belavkin, G. M. D'Ariano, and M. Raginsky, Operational distance and fidelity for quantum channels, *J. Math. Phys.* **46**, 062106 (2005).
- [4] I. L. Chuang and M. A. Nielsen, Prescription for experimental determination of the dynamics of a quantum black box, *J. Mod. Opt.* **44**, 2455 (1997).
- [5] S. Aaronson, Shadow tomography of quantum states, in *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2018, edited by I. Diakonikolas, D. Kempe, and M. Henzinger (Association for Computing Machinery, New York, 2018), pp. 325–338.
- [6] H.-Y. Huang, R. Kueng, and J. Preskill, Predicting many properties of a quantum system from very few measurements, *Nat. Phys.* **16**, 1050 (2020).
- [7] S. Chen, W. Yu, P. Zeng, and S. T. Flammia, Robust shadow estimation, *PRX Quantum* **2**, 030348 (2021).
- [8] C. Hadfield, S. Bravyi, R. Raymond, and A. Mezzacapo, Measurements of quantum Hamiltonians with locally-biased classical shadows, *Commun. Math. Phys.* **391**, 951 (2022).
- [9] H.-Y. Huang, R. Kueng, and J. Preskill, Efficient estimation of Pauli observables by derandomization, *Phys. Rev. Lett.* **127**, 030503 (2021).
- [10] A. Elben, S. T. Flammia, H.-Y. Huang, R. Kueng, J. Preskill, B. Vermersch, and P. Zoller, The randomized measurement toolbox, *Nat. Rev. Phys.* **5**, 9 (2023).
- [11] R. Levy, D. Luo, and B. K. Clark, Classical shadows for quantum process tomography on near-term quantum computers, *Phys. Rev. Res.* **6**, 013029 (2024).
- [12] Y. Wu, W.-S. Bao, S. Cao, F. Chen, M.-C. Chen, X. Chen, T.-H. Chung, H. Deng, Y. Du, D. Fan, M. Gong, C. Guo, C. Guo, S. Guo, L. Han, L. Hong, H.-L. Huang, Y.-H. Huo, L. Li, N. Li *et al.*, Strong quantum computational advantage using a superconducting quantum processor, *Phys. Rev. Lett.* **127**, 180501 (2021).
- [13] F. Arute, K. Arya, R. Babbush, D. Bacon, J. Bardin, R. Barends, R. Biswas, S. Boixo, F. Brandao, D. Buell, B. Burkett, Y. Chen, Z. Chen, B. Chiaro, R. Collins, W. Courtney, A. Dunsworth, E. Farhi, B. Foxen, and J. Martinis, Quantum supremacy using a programmable superconducting processor, *Nature (London)* **574**, 505 (2019).
- [14] J. Helsen, I. Roth, E. Onorati, A. H. Werner, and J. Eisert, General framework for randomized benchmarking, *PRX Quantum* **3**, 020357 (2022).

- [15] E. Knill, D. Leibfried, R. Reichle, J. Britton, R. B. Blakestad, J. D. Jost, C. Langer, R. Ozeri, S. Seidelin, and D. J. Wineland, Randomized benchmarking of quantum gates, *Phys. Rev. A* **77**, 012307 (2008).
- [16] J. Helsen, M. Ioannou, J. Kitzinger, E. Onorati, A. Werner, J. Eisert, and I. Roth, Estimating gate-set properties from random sequences, *Nat. Commun.* **14**, 5039 (2023).
- [17] R. Fan, P. Zhang, H. Shen, and H. Zhai, Out-of-time-order correlation for many-body localization, *Sci. Bull.* **62**, 707 (2017).
- [18] C.-F. A. Chen, A. Lucas, and C. Yin, Speed limits and locality in many-body quantum dynamics, *Rep. Prog. Phys.* **86**, 116001 (2023).
- [19] H. Shen, P. Zhang, Y.-Z. You, and H. Zhai, Information scrambling in quantum neural networks, *Phys. Rev. Lett.* **124**, 200504 (2020).
- [20] L. Leone, S. F. E. Oliviero, Y. Zhou, and A. Hamma, Quantum chaos is quantum, *Quantum* **5**, 453 (2021).
- [21] B. Vermersch, A. Elben, L. Sieberer, N. Yao, and P. Zoller, Probing scrambling using statistical correlations between randomized measurements, *Phys. Rev. X* **9**, 021061 (2019).
- [22] B. Sundar, A. Elben, L. K. Joshi, and T. V. Zache, Proposal for measuring out-of-time-ordered correlators at finite temperature with coupled spin chains, *New J. Phys.* **24**, 023037 (2022).
- [23] A. M. Green, A. Elben, C. H. Alderete, L. K. Joshi, N. H. Nguyen, T. V. Zache, Y. Zhu, B. Sundar, and N. M. Linke, Experimental measurement of out-of-time-ordered correlators at finite temperature, *Phys. Rev. Lett.* **128**, 140601 (2022).
- [24] J.-H. Wang, T.-Q. Cai, X.-Y. Han, Y.-W. Ma, Z.-L. Wang, Z.-H. Bao, Y. Li, H.-Y. Wang, H.-Y. Zhang, L.-Y. Sun *et al.*, Verifying quantum information scrambling dynamics in a fully controllable superconducting quantum simulator, [arXiv:2112.11204](https://arxiv.org/abs/2112.11204).
- [25] B. Yoshida and N. Y. Yao, Disentangling scrambling and decoherence via quantum teleportation, *Phys. Rev. X* **9**, 011006 (2019).
- [26] X. Mi, P. Roushan, C. Quintana, S. Mandrà, J. Marshall, C. Neill, F. Arute, K. Arya, J. Atalaya, R. Babbush, J. C. Bardin, R. Barends, J. Basso, A. Bengtsson, S. Boixo, A. Bourassa, M. Broughton, B. B. Buckley, D. A. Buell, B. Burkett *et al.*, Information scrambling in quantum circuits, *Science* **374**, 1479 (2021).
- [27] R. J. Garcia, Y. Zhou, and A. Jaffe, Quantum scrambling with classical shadows, *Phys. Rev. Res.* **3**, 033155 (2021).
- [28] L. Leone, S. F. E. Oliviero, and A. Hamma, Nonstabilizerness determining the hardness of direct fidelity estimation, *Phys. Rev. A* **107**, 022429 (2023).
- [29] S. Xu and B. Swingle, Locality, quantum fluctuations, and scrambling, *Phys. Rev. X* **9**, 031048 (2019).
- [30] T. Schuster, J. Haferkamp, and H.-Y. Huang, Random unitaries in extremely low depth, [arXiv:2407.07754](https://arxiv.org/abs/2407.07754).
- [31] J. Helsen, J. J. Wallman, and S. Wehner, Representations of the multi-qubit Clifford group, *J. Math. Phys.* **59**, 072201 (2018).
- [32] C. Dankert, R. Cleve, J. Emerson, and E. Livine, Exact and approximate unitary 2-designs and their application to fidelity estimation, *Phys. Rev. A* **80**, 012304 (2009).
- [33] D. Porras and J. I. Cirac, Effective quantum spin systems with trapped ions, *Phys. Rev. Lett.* **92**, 207901 (2004).
- [34] H. Zhu, Multiqubit clifford groups are unitary 3-designs, *Phys. Rev. A* **96**, 062336 (2017).