Quantifying information scrambling via classical shadow tomography on programmable quantum simulators

Max McGinley [1], Sebastian Leontica [1], Samuel J. Garratt,[1,2] Jovan Jovanovic,[1] and Steven H. Simon[1]

[1]*Rudolf Peierls Centre for Theoretical Physics, Clarendon Laboratory,*
Oxford University, Parks Road, Oxford OX1 3PU, United Kingdom
[2]*Department of Physics, University of California, Berkeley, Berkeley, California 94720, USA*

We develop techniques to probe the dynamics of quantum information and implement them experimentally on an IBM superconducting quantum processor. Our protocols adapt shadow tomography for the study of time-evolution channels rather than of quantum states and rely only on single-qubit operations and measurements. We identify two unambiguous signatures of quantum information scrambling, neither of which can be mimicked by dissipative processes, and relate these to many-body teleportation. By realizing quantum chaotic dynamics in experiment, we measure both signatures and support our results with numerical simulations of the quantum system. We additionally investigate operator growth under this dynamics and observe behavior characteristic of quantum chaos. As our methods require only a single quantum state at a time, they can be readily applied on a wide variety of quantum simulators.

I. INTRODUCTION

Scrambling is fundamental to our current understanding of many-body quantum dynamics in fields ranging from thermalization and chaos [1–4] to black holes [5–7]. This is the process by which initially local information, such as charge imbalance in a solid, becomes hidden in increasingly nonlocal degrees of freedom under unitary time evolution. Scrambling accounts for both the fate of information falling into black holes [5,8] and the apparent paradox of equilibration under unitary dynamics: Information about the initial state is not truly lost, but rather becomes inaccessible when one can only measure local observables, as is the case in traditional experimental settings.

Today, the experimental settings we have access to offer a much higher degree of control and programmability than those that were available when these questions were first addressed. New kinds of quantum devices can be constructed by assembling qubits that are individually addressable, such as those made from trapped ions [9–12], superconducting circuits [13–17], or Rydberg atoms [18–23]. Such noisy intermediate-scale quantum (NISQ) devices [24] allow a wider range of interactions to be synthesized and crucially permit measurements of highly nonlocal observables, making the distinction between nonunitary information loss and unitary information scrambling more than a purely academic one. As well as providing further motivation for theoretical work on quantum chaos and scrambling, these technological developments present an opportunity for complementary experimental studies, which promise to be of increasing utility as the size and complexity of the systems continue to grow beyond what can be simulated classically [25,26].

A variety of experimental protocols to probe quantum chaos have already been put forward and implemented, with early approaches based on measuring the growth of quantum entanglement. For example, if two copies of the system can be prepared simultaneously, then certain quantifiers of entanglement can be extracted from joint measurements on the two copies [27–31]. More recently, focus has shifted towards probing scrambling rather than entanglement growth, primarily via so-called out-of-time-order correlators (OTOCs) [7,32–35], which can be measured when the dynamics can be time reversed [36–40]. However, the link between OTOC decay and scrambling is predicated on the assumption that the dynamics is unitary [41]; this is invariably not the case in NISQ devices, which are by definition noisy. Moreover, with system sizes being somewhat limited at present, protocols that are qubit efficient (i.e., not requiring multiple copies of the system at once) will be required to make progress in the near term.

Emphasizing its practical implementation in an IBM superconducting quantum computer, in this work we show how scrambling can be quantified in NISQ devices using only single-qubit manipulations and individual copies of a quantum state at a time. To achieve this, we first generalize the technique of shadow tomography [42] to study dynamics. We then prove that certain well-established physical quantities are (i) accessible using this technique and (ii) provide unambiguous signatures of scrambling. Crucially, the signatures that we identify remain meaningful even when the system's dynamics is nonunitary; this allows us to verifiably detect scrambling on a real noisy quantum device.

The quantities that we identify satisfying the above two criteria are related to operator-space entanglement (OE), also known as entanglement in time [43–46]. While entanglement quantifies quantum correlations between degrees of freedom at one instant in time, OE pertains to correlations that are conveyed across time, which is of direct relevance to scrambling.

This has proved to be an extremely useful tool in analytical and numerical studies of chaotic quantum dynamics [45,47–54], allowing one to construct measures of chaos in a dynamical, manifestly state-independent way.

Here we establish a link between OE and the ability of a system to transmit information from one qubit to another via a process known as many-body teleportation, or Hayden-Preskill teleportation after the authors of Ref. [5]. This process was originally considered in the context of the black hole information paradox [8] and is now a central part of the theory of scrambling. We put forward two OE-based quantities [Eqs. (2) and (3)] and show that each can be related to the fidelity of Hayden-Preskill teleportation. In particular, we argue that both quantities have a threshold value which when exceeded gives a guarantee that the quantum communication capacity from one qubit to another is nonzero, i.e., quantum states can be reliably transmitted at a finite rate using the quantum system as a communication channel, even when the dynamics is nonunitary.

Beyond establishing these quantities as meaningful measures of scrambling, we demonstrate their practical utility by showing that both are directly measurable in experiment. The scheme we introduce allows one to measure the necessary information-theoretic quantities with minimal experimental overhead. This is made possible by extending ideas originally developed to measure entanglement in an instantaneous state. In that context, it has been demonstrated that measurements of the state in randomly selected bases can be used to extract certain entanglement measures [42,55,56], without requiring joint access to multiple copies of the state per experiment. To generalize from state entanglement to OE, we propose to prepare initial states in random bases, which are then time evolved under the dynamics of interest, before being measured in random bases (see Fig. 2). By postprocessing the classical data generated by this sequence of operations in a way analogous to that proposed in Ref. [42], we are able to construct estimators of the quantities in question. We do so explicitly using data from an IBM quantum computer, giving us access to spatially resolved measures of information delocalization, revealing the light-cone structure in the system's dynamics.

In addition to these probes of many-body teleportation, our protocol can be used to obtain a fine-grained description of operator spreading [34,57–59]. Specifically, shadow tomography of the dynamics gives us access to certain combinations of the operator spreading coefficients studied in Ref. [58], which gives a complementary perspective on scrambling.

Other quantities related to operator entanglement, namely, out-of-time-order correlators [7,32–35], have been measured in previous experiments [36–40] and indeed are in principle measurable using shadow tomography and related methods [60,61]. However, these cannot be used as an unambiguous diagnostic of scrambling, since dissipation and miscalibrations can give rise to the same signal as that of a true scrambler [41]. In contrast, the quantities (2) and (3) measured here constitute a positive verifiable signature of scrambling, which cannot be mimicked by noise. We note that related signatures of teleportation have been observed before using multiple copies of the system evolving in a coordinated fashion [62,63]. A key innovation in our work is to quantify the fidelity of teleportation without actually performing teleportation. As a
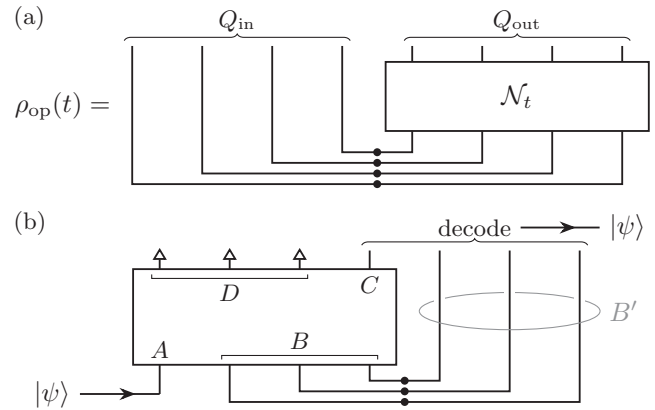

FIG. 1. (a) Representation of the operator state $\rho_{\mathrm{op}}(t)$ [Eq. (1)]. Each qubit in Q_{out} is prepared in a maximally entangled state (black dots) with the corresponding qubit Q_{in}, before being time evolved under the channel $\mathcal{N}_t$. (b) Illustration of the Hayden-Preskill protocol [5]. An unknown quantum state $|\psi\rangle$ is used as an input to a small subregion A, while the remaining qubits B are prepared in a maximally entangled state with a set of ancillas B' (circled). If the channel is perfectly scrambling then $|\psi\rangle$ can be reconstructed using the ancillas combined with a subset of output qubits C of the same size as A, regardless of which qubits are in C (qubits in D are discarded). Formally, the final state of the ancillas combined with the outputs C depends on the input state to A through the channel $\mathcal{N}_t^{A \to B'C}$ (see the text for details).

consequence, our method can probe scrambling with half as many qubits and without needing to match the time evolution between two separate systems, which may not be possible when the dynamics is not known *a priori*.

As well as superconducting qubits, the protocol we use here is implementable using presently available techniques in a variety of other platforms including those based on Rydberg atom arrays [18,20–22], trapped ions [11,12], and photonics [26,64–66]. We compare the protocol to previous approaches used to diagnose scrambling and discuss the tradeoffs between sample efficiency, verifiability, and the required degree of experimental control.

This paper is organized as follows. In Sec. II A we introduce the concept of operator-space entanglement as well
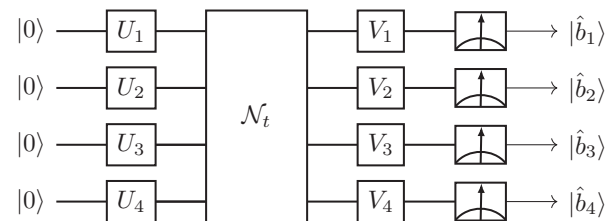

FIG. 2. Illustration of experimental protocol to measure operator-space entanglement of a quantum channel $\mathcal{N}_t$ in a system with $N = 4$ qubits. The single-qubit unitaries U_j and V_j are drawn independently at random from the discrete gate sets described in the text. Once the measurement outcomes $\hat{b}_j$ are known, one can construct a snapshot of the doubled state $\rho_{\mathrm{op}}(t)$ using Eq. (7) and then repeat M times with different unitaries.

as the Hayden-Preskill protocol for many-body teleportation [5] and describe how the two are related. We then introduce the key quantities (2) and (3) that we will use to quantify Hayden-Preskill teleportation in Sec. II B as well as showing how OE allows one to track the growth of operators under Heisenberg time evolution. Section III describes our shadow tomographic protocol that can be used to estimate the above quantities. Results from implementing this protocol on an IBM superconducting quantum processor are given in Sec. IV. We discuss our results and present our conclusions in Sec. V.

II. PROBING SCRAMBLING USING OPERATOR-SPACE ENTANGLEMENT

A. Operator-space entanglement and the Hayden-Preskill protocol

The evolution of a quantum system Q with Hilbert space $\mathcal{H}_Q$ from time 0 to time t can be described by a channel $\mathcal{N}_t$ such that the density matrix evolves as $\rho^Q(t) = \mathcal{N}_t[\rho^Q(0)]$. The usual notion of entanglement in a state can be generalized to channels, which is known as operator-space entanglement. Formally, this is done by reinterpreting $\mathcal{N}_t$ as a state on a doubled Hilbert space [67,68], on which conventional entanglement measures can be defined. This is perhaps most simply understood when the dynamics is unitary $\mathcal{N}_t[\rho^Q] = U_t \rho^Q U_t^\dagger$, as detailed in Ref. [45]. Fixing a basis of product states $\{|a\rangle\}$ for Q, a pure doubled state (existing in operator space) is constructed as $|U_t\rangle_{\mathrm{op}} = |\mathcal{H}_Q|^{-1/2} \sum_{ab} \langle b|U_t|a\rangle \, |a\rangle_{\mathrm{in}} \otimes |b\rangle_{\mathrm{out}}$, where $|\mathcal{H}_Q|$ is the Hilbert space dimension and the in and out labels refer to the inputs and outputs of the unitary, respectively. In words, the components of $|U_t\rangle_{\mathrm{op}}$ are the $|\mathcal{H}_Q|^2$ matrix elements of the unitary U_t. From here onward we specialize to N-qubit systems, so $|\mathcal{H}_Q| = 2^N$.

This construction has an alternative interpretation: $|U_t\rangle_{\mathrm{op}}$ is the state that results from evolving a maximally entangled state $|\Phi\rangle = 2^{-N/2} \sum_a |a\rangle_{\mathrm{in}} \otimes |a\rangle_{\mathrm{out}}$ under the unitary $\mathbb{I}_{\mathrm{in}} \otimes U_t$, i.e., one half of the maximally entangled pair is evolved under U_t. This also makes it clear how to generalize to nonunitary evolutions: $|U_t\rangle_{\mathrm{op}}$ is replaced by a mixed state

$$\rho_{\mathrm{op}}(t) = (\mathrm{id}_{\mathrm{in}} \otimes \mathcal{N}_t)[|\Phi\rangle \langle \Phi|]. \tag{1}$$

This construction is illustrated in Fig. 1(a). We use the more generally applicable density matrix $\rho_{\mathrm{op}}(t)$, rather than the pure state $|U_t\rangle_{\mathrm{op}}$, in the following. Note that correlation functions with respect to the doubled state $\mathrm{Tr}[(O_{\mathrm{in}} \otimes O_{\mathrm{out}})\rho_{\mathrm{op}}(t)]$ map to infinite-temperature two-time correlation functions $2^{-N}\mathrm{Tr}[O_{\mathrm{in}}^T O_{\mathrm{out}}(t)]$ (where time evolution of operators in the Heisenberg picture is given by $O_{\mathrm{out}}(t) = \mathcal{N}_t^\dagger[O_{\mathrm{out}}]$ and the transpose is taken with respect to the basis $\{|a\rangle\}$).

Evidently, at $t = 0$ ($\mathcal{N}_{t=0} = \mathrm{id}$) a given input qubit with index j_{in} is maximally entangled with the corresponding output qubit $j_{\mathrm{out}} = j_{\mathrm{in}}$ only. This reflects the trivial observation that information is perfectly transmitted from j_{in} to $j_{\mathrm{out}} = j_{\mathrm{in}}$ under $\mathcal{N}_{t=0}$. If $\mathcal{N}_t$ exhibits scrambling, then we expect that locally encoded information will begin to spread out as the output qubits evolve such that j_{in} becomes entangled with many other output qubits. At late times, one will no longer be able to extract these correlations from any small output region

C; instead, the information about the initial state of a given qubit will be encoded across many output qubits.

This intuition can be quantified in terms of particular measures of operator-space entanglement. These are constructed by evaluating familiar quantities associated with state entanglement on $\rho_{\mathrm{op}}(t)$. In the doubled space, one can divide the input qubits into A and its complement B, and the outputs into C and its complement D. (A and C need not correspond to the same physical qubits.) Reduced density matrices can then be formed, e.g., $\rho^{AC}(t) = \mathrm{Tr}_{B \cup D} \rho_{\mathrm{op}}(t)$. Two important information-theoretic quantities are the von Neumann entanglement entropy $S(AC) = -\mathrm{Tr}\rho^{AC}(t) \log_2 \rho^{AC}(t)$ and the mutual information $I(A : C) = S(A) + S(C) - S(AC)$ (all logarithms are base 2 and we leave the t dependence of entropies and mutual information implicit). The mutual information quantifies the degree to which the initial state of qubits in A is correlated with the final state of qubits in C (this includes both classical and quantum correlations). Indeed, $I(A : C)$ is closely related to the capacity of the channel for classical communication from a sender A to a receiver C [69,70].

Given that the reduced density matrices $\rho^{AC}(t)$ will typically be highly mixed, it is also useful to examine quantities that have been devised to probe mixed state entanglement. The logarithmic negativity $E_{A:C} := \log_2 \mathrm{Tr}|\rho^{AC}(t)^{T_A}|$ (where T_A denotes a partial transpose on A and $|\mathcal{O}| := \sqrt{\mathcal{O}^\dagger \mathcal{O}}$ for operators $\mathcal{O}$) is useful for this purpose: When applied to a bipartite state it can be used to bound the distillable entanglement between A and C [71,72], which unlike mutual information excludes classical correlations. Here we will consider the operator-space generalization of negativities, which have been connected to scrambling in the context of random unitary circuits and holographic channels [73].

As argued by the authors of Ref. [45], for unitary chaotic channels the correlations between regions A and C of size $O(1)$ will be small, whereas $I(A : CD)$ will be maximal, indicating that the input state A can only be reconstructed if one has access to all the outputs CD. They propose the tripartite information $I_3(A : C : D) = I(A : C) + I(A : D) - I(A : CD)$ as a diagnostic of scrambling (for scramblers I_3 is large and negative), illustrating one way in which operator-space entanglement measures can be used to detect scrambling.

A complementary way to diagnose scrambling is to quantify correlations between A and BC that are present in $\rho_{\mathrm{op}}(t)$, where again A and C are of size $O(1)$. This approach is related to the Hayden-Preskill teleportation problem [5], a thought experiment that was initially devised to understand the fate of information in black holes. There one asks if it is possible to recover the initial state of a small set of qubits A using a set of ancillas B' that are initially maximally entangled with B, combined with a subset of output qubits C [see Fig. 1(b)]. If $\mathcal{N}_t$ is scrambling, then the initial state of A becomes nonlocally encoded across the entire system. When this occurs, teleportation can be achieved (i.e., the initial state of A can be recovered from $B'C$) regardless of which qubits are chosen in C, as long as $|C| \geqslant |A|$ [62].

Intuitively, we expect that for teleportation to be successful, there must be strong correlations between A and BC in the state $\rho_{\mathrm{op}}(t)$. This can in principle be diagnosed using the

quantities introduced above, namely, $I(A:BC)$ and $E_{A:BC}$. More formally, we can capture the dependence of the final state of $B'C$ on the initial state A using the channel $\mathcal{N}_t^{A\to B'C}[\rho^A] = \mathrm{Tr}_D\{(\mathcal{N}_t \otimes \mathrm{id}_{B'})[\rho^A \otimes \Phi_{BB'}]\}$. The fidelity of teleportation in the Hayden-Preskill protocol is then determined by the potential for information transmission through $\mathcal{N}_t^{A\to B'C}$, which can be quantified in an information-theoretic way using an appropriate channel capacity [74]. As an example, the classical capacity of $\mathcal{N}_t^{A\to B'C}$ is closely related to $I(A:BC)$ [69,70]. Similarly, the quantum channel capacity (the maximum rate at which quantum states can be reliably transmitted using multiple applications of the channel) can be bounded by $E_{A:BC}$ [75]. This illustrates the connection between information transmission in the Hayden-Preskill protocol and the degree of correlations between A and BC in the operator state $\rho_{\mathrm{op}}(t)$.

The experiment of Ref. [62] provided an explicit demonstration of scrambling by executing a particular decoding procedure for the Hayden-Preskill protocol. This requires one to construct a doubled state and manipulate the ancillas B'. In contrast, in this paper we quantify the correlations between A and BC without ever performing the teleportation explicitly. This avoids us having to construct a doubled state or execute a decoding procedure.

B. Rényi measures of scrambling and operator growth

While the von Neumann entropy and quantities derived thereof have strong information-theoretic significance, they are not directly measurable in experiments without recourse to full tomography of $\rho_{\mathrm{op}}(t)$, which is computationally expensive [76]. This is due to the need to take the operator logarithm of ρ. Instead, one can generalize to Rényi entropies $S^{(m)}(AC) := (1-m)^{-1} \log_2 \mathrm{Tr}\{[\rho^{AC}(t)]^m\}$ $(m = 2, 3, \dots)$, which unlike $S(AC)$ only depend on integer moments of the density matrix and hence can be computed in terms of mth moments of correlation functions of $\rho^{AC}(t)$. This observation forms the basis of a number of protocols which use randomized measurements to extract the Rényi entropies of an instantaneous state [42,55], as well as integer moments of the density matrix after partial transposition [77]. Later, we will employ similar arguments to show that the analogous quantities in operator space can also be directly measured. Before doing so, we first discuss how these quantities can be used to probe quantum chaotic dynamics and information scrambling, making use of the insight described in the preceding section.

We have argued how $I(A:BC)$ can be related to the fidelity of the Hayden-Preskill protocol. A natural generalization of $I(A:BC)$ that is constructed in terms of integer moments of ρ_{op} is the Rényi mutual information

$$I^{(m)}(A:BC) := S^{(m)}(A) + S^{(m)}(BC) - S^{(m)}(ABC). \quad (2)$$

When evaluated on arbitrary states this simple generalization of the mutual information does not satisfy all the same properties as $I(A:BC)$, including non-negativity [78–80]. However, in Appendix A we show that when evaluated on operator states (1) (for which the reduced density matrix on A is maximally mixed), $I^{(m)}(A:BC)$ is non-negative [46] and equal to zero if and only if A and BC are uncorrelated, as one would desire for

any measure of correlation. Additionally, for $m = 2$ the Rényi mutual information is related to the recovery fidelity F for the decoding protocol used in Ref. [62] by $F = 2^{I^{(2)}(A:BC)-2|A|}$ [41] and can also be expressed in terms of particular sums of two-point correlation functions or OTOCs [45,46].

Given the above, we expect that the quantity (2) will be sensitive to the temporal correlations that are conveyed by channels that exhibit scrambling. Moreover, while mutual information captures classical and quantum correlations on an equal footing, one can still use $I^{(m)}(A:BC)$ to detect the transmission of purely quantum information. Specifically, we argue that the channel $\mathcal{N}_t^{A\to B'C}$, which describes the Hayden-Preskill setup, must have a nonzero quantum communication capacity if $I^{(m)}(A:BC)$ exceeds the threshold value of $|A|$, which is the maximum value that can be obtained in a classical system. The full proof of this statement is given in Appendix A. In brief, we show that violation of the classical limit can only occur if there is entanglement between A and BC in the operator state $\rho_{\mathrm{op}}(t)$. Given multiple uses of the channel, one can distill this entanglement into Einstein-Podolsky-Rosen (EPR) pairs, which can then be used for noiseless quantum communication. This confirms that $\mathcal{N}_t^{A\to B'C}$ can in principle be used to reliably transmit quantum information and thus the quantum capacity is nonzero. Note that the converse is not necessarily true, i.e., there exist channels for which the quantum capacity is nonzero but $I^{(m)}(A:BC) \leqslant |A|$.

We can also consider quantities related to negativity that only involve integer moments of the density matrix. Let us first define moments of the partially transposed operator state $p_{m,X:Y} := \mathrm{Tr}\{[\rho^{XY}(t)^{T_X}]^m\}$, where X and Y are nonoverlapping sets of input and output qubits and again T_X denotes a partial transpose on X. We will consider the quantity

$$R_{A:BC} := \frac{p_{2,A:BC}^2}{p_{3,A:BC}}. \quad (3)$$

This particular ratio was proposed as a measure of mixed state entanglement in Ref. [77], where it was shown that bipartite states ρ^{AB} satisfying $R_{A:B} > 1$ must be entangled. In Appendix A we argue that $R_{A:BC} > 1$ is a sufficient (but not necessary) condition for the quantum communication capacity of $\mathcal{N}_t^{A\to B'C}$ to be nonzero provided A is a single qubit (which is the case throughout this paper).

The above arguments demonstrate how the Rényi generalizations of mutual information and negativity can be related to the Hayden-Preskill teleportation fidelity. A complementary way to probe aspects of chaos in quantum dynamics is to consider the time evolution of operators in the Heisenberg picture $\mathcal{O}(t) = \mathcal{N}_t^\dagger[\mathcal{O}]$ [57,58]. Operator-space Rényi entropies for $m = 2$ (equivalently, operator-space purities $\mathrm{Tr}[\rho^{AC}(t)^2] \equiv 2^{-S^{(2)}(AC)}$) can be related to the structure of operator growth. To see this, let us use Pauli strings $\boldsymbol{\sigma}^\mu = \bigotimes_j \sigma_j^{\mu_j}$ as a basis of operators, where $\mu = (\mu_1, \dots, \mu_N)$ and $\mu_j \in \{I, X, Y, Z\}$. Adopting the notation of Ref. [58], operator spreading coefficients $c^{\mu\nu}(t)$ can then be defined via an expansion of time-evolved Pauli strings $\boldsymbol{\sigma}^\mu(t) = \mathcal{N}_t^\dagger[\boldsymbol{\sigma}^\mu]$, namely, $\boldsymbol{\sigma}^\mu(t) = \sum_\nu c^{\mu\nu}(t)\boldsymbol{\sigma}^\nu$. It is straightforward to show that operator-space purity can be expressed succinctly in terms of operator spreading

coefficients as

$$\text{Tr}[\rho^{AC}(t)^2] = \frac{1}{2^{|A|+|C|}} \sum_{\nu \in A} \sum_{\mu \in C} |c^{\mu\nu}(t)|^2, \qquad (4)$$

where the sums are over Pauli strings ν and μ that act as identity on qubits outside of A and C, respectively. In words, we identify operator-space purity as the norm of the part of the evolved operator $\boldsymbol{\sigma}^\mu(t)$ that has support on A, averaged over all initial operators $\boldsymbol{\sigma}^\mu$ with support on C.

Equation (4) clarifies how operator purities encode the spatial structure of operator spreading. One concise way to represent this information is in terms of the k-locality of the evolved operator $\boldsymbol{\sigma}^\mu(t)$, i.e., one can ask what proportion of the Pauli strings that make up $\boldsymbol{\sigma}^\mu(t)$ act nontrivially on at most k qubits. Intuitively, local operators with support on a small number of qubits will grow under chaotic time evolution, leading to more weight on operators that have a wider support. This contrasts with integrable systems, where $\boldsymbol{\sigma}^\mu(t)$ spreads out in space without becoming more complex in terms of k-locality.

A natural way to measure k-locality of the evolved operator $\boldsymbol{\sigma}^\mu(t)$ is to compute the norm of the part of the operator that is made up of Pauli strings acting on exactly k qubits

$$D_k^\mu(t) := \sum_{\nu:|\boldsymbol{\sigma}^\nu|=k} |c^{\mu\nu}(t)|^2, \qquad (5)$$

where we use $|\boldsymbol{\sigma}^\nu|$ to denote the number of nonidentity factors in the string $\boldsymbol{\sigma}^\nu$. If one takes an average of $D_k^\mu(t)$ over all nonidentity Pauli strings μ with support in some region C, the resulting quantity can be expressed in terms of operator purities

$$D_k^C(t) := \frac{1}{2^{|C|}-1} \sum_{\mu \in C; \mu \neq I^{\times N}} D_k^\mu(t), \quad k \geqslant 1$$

$$= \frac{2^{|C|}(-1)^k}{2^{|C|}-1} \sum_{\substack{A \subseteq S \\ |A| \leqslant k}} (-2)^{|A|} \binom{N-|A|}{N-k} \text{Tr}[\rho^{AC}(t)^2]. \quad (6)$$

We prove the second equation in Appendix B. The above quantity allows one to track how operators initially located within C increase in complexity (in the sense of k-locality) with time. Later, we will use $D_k^C(t)$ as a means to quantify this aspect of operator growth on a quantum device.

In the following section, we demonstrate that the quantities described above, which depend only on integer moments of the operator state $\rho_{\text{op}}(t)$, can be directly measured in experiment without using full tomography. Moreover, this can be done without ever explicitly constructing the doubled state, which would require simultaneous access to identical copies of the system.

III. SHADOW TOMOGRAPHIC MEASUREMENT OF OPERATOR-SPACE ENTANGLEMENT

The method we use to measure operator-space Rényi entropies is based on classical shadow tomography [42]. There one performs projective measurements in different randomly selected bases on a target state ρ, each of which gives a particular snapshot of ρ. The ensemble of snapshots (known

as the shadow of ρ) has an efficient classical representation, which allows one to calculate estimators of expectation values $\text{Tr}[\mathcal{O}\rho]$ and nonlinear moments $\text{Tr}[A\rho^{\otimes m}]$ through classical processing of shadow data.

Here we propose to build up a shadow of the doubled state $\rho_{\text{op}}(t)$ by preparing random states, evolving them under $\mathcal{N}_t$, and performing measurements in independently chosen random bases. For our purposes, the random states and bases will be related to the computational basis by single-qubit rotations, since these can be implemented accurately on current devices; however, generalizations to global rotations are also possible [42,81].

The specific protocol is illustrated in Fig. 2. Output rotations V_j applied immediately prior to measurement are sampled independently from a uniform distribution over the discrete set of gates $\{\mathbb{I}, H_X, H_Y\}$, where $H_{X,Y}$ are X and Y Hadamard gates. This effectively implements one of the three possible Pauli measurements for each qubit. The gates U_j applied prior to time evolution are chosen such that the distribution of initial input states $U_j |0\rangle$ is uniform over the six states $\{|\pm_\sigma\rangle : \sigma = X, Y, Z\}$, where $|+_\sigma\rangle$ ($|-_\sigma\rangle$) is the eigenstate of the Pauli operator σ with eigenvalue $+1$ (-1). A total of M runs are performed and for now we assume that a new set of independent gates are generated for each run.

The data associated with a particular run are the gates U_j and V_j, along with the measurement outcomes $\hat{b}_j \in \{0, 1\}$. These can be used to construct a snapshot of $\rho_{\text{op}}(t)$ (we use a hat to denote random variables, allowing us to distinguish this estimator from the true operator state),

$$\hat{\rho}_{\text{op}}(t) = \bigotimes_{j=1}^N \left(3U_j^T |0\rangle \langle 0| U_j^* - \mathbb{I}\right)_{\text{in}}$$

$$\times \otimes \bigotimes_{j=1}^N (3V_j |\hat{b}_j\rangle \langle \hat{b}_j| V_j^\dagger - \mathbb{I})_{\text{out}}. \qquad (7)$$

Using the arguments of Ref. [42], along with the definition of $\rho_{\text{op}}(t)$ and the property of the maximally entangled state $(\mathcal{O} \otimes \mathbb{I}) |\Phi\rangle = (\mathbb{I} \otimes \mathcal{O}^T) |\Phi\rangle$, one can show that the above is an unbiased estimator of $\rho_{\text{op}}(t)$, i.e., $\mathbb{E}[\hat{\rho}_{\text{op}}(t)] = \rho_{\text{op}}(t)$, where the expectation value is over both random unitaries U_j and V_j and measurement outcomes $\hat{b}_j$. [See Appendix C. Equation (7) is consistent with other similar proposals that have appeared recently [82,83].] A different snapshot is obtained from each of the M runs and we write the snapshot obtained from the rth run as $\hat{\rho}_{\text{op}}^{(r)}(t)$.

For each $\hat{\rho}_{\text{op}}^{(r)}(t)$, an independent unbiased estimator of a given correlation function $\text{Tr}[\mathcal{O}_{\text{in}} \mathcal{O}_{\text{out}}(t)]$ can be constructed by computing $\text{Tr}[(\mathcal{O}_{\text{in}} \otimes \mathcal{O}_{\text{out}}) \hat{\rho}_{\text{op}}^{(r)}(t)]$ on a classical computer. For sufficiently large M, the average over all estimators gives an accurate prediction of the correlation function. Estimators for nonlinear functionals, such as the moments $p_{m,:AC} = \text{Tr}[\rho^{AC}(t)^m]$ appearing in the Rényi entropies, can be constructed using so-called U statistics [84], as one does in conventional shadow tomography. For instance, to estimate $p_{2,:AC}$, one can average $\text{Tr}[\hat{\rho}^{AC,(r_1)}(t) \hat{\rho}^{AC,(r_2)}(t)]$ over all $M(M-1)$ ordered pairs of independent snapshots $r_1 \neq r_2$, where $\rho^{AC,(r)}(t) := \text{Tr}_{B \cup D} \hat{\rho}_{\text{op}}^{(r)}(t)$. The snapshot (7) can also be partially transposed beforehand to obtain $p_{m,A:C}$.
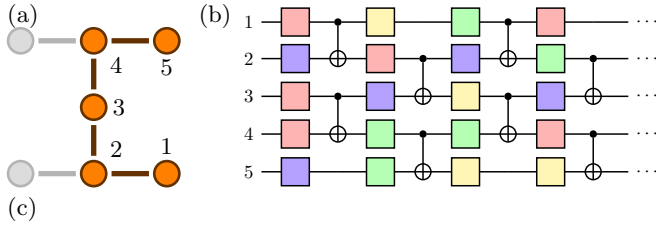
FIG. 3. (a) Qubit layout and connectivity of ibm_lagos. Purple circles represent the five qubits used for the experiments detailed in the text. (b) Circuit design for the chaotic unitary $\mathcal{N}_t$, with $t = 4$ time steps shown. Each single-qubit gate (colored boxes) is independently sampled from the four gates $W_{1,\ldots,4}$ (see the text for details).

Conveniently, the same set of shadow data can be used to obtain multiple quantities simply by postprocessing in different ways.

The size of the statistical errors that arise from this process will depend on the particular quantity being estimated, the channel $\mathcal{N}_t$ in question, and the sample count M. Worst-case upper bounds on the number of samples M_ϵ required to achieve an error ϵ in state shadow tomography have been derived in Refs. [42,77], and these can be carried over to the present setting, at least for single-qubit rotations. For the moments $p_{m,:AC} = \text{Tr}[\rho^{AC}(t)^m]$ (with or without partial transposition) in the small-ϵ limit, one has $M_\epsilon \leqslant O(2^{|AC|}/\epsilon^2)$. In the Supplemental Material [85] we argue that when $\rho^{AC}(t)$ is highly mixed (which is common for operator-space states), a potentially tighter upper bound of $O(2^{|AC| \log_2 3 - S^{(\infty)}(AC)}/\epsilon^2)$ applies, where $S^{(\infty)}(AC) = -\log_2 \max \text{eig} \, \rho^{AC}(t)$ is the max entropy. While this is exponential in the number of qubits in AC, the scaling is highly favorable over the $O(2^{|AC|}\text{rank}(\rho_{AC})^2/\epsilon^2 \log_2(1/\epsilon))$ number of runs required for full tomography using the same resources (i.e., only single-qubit rotations) [87]. In general, while these bounds are expected to have the correct scaling behavior, the prefactors involved are typically not tight [42].

IV. SIMULATING AND DETECTING QUANTUM CHAOS

We now present results of simulations of quantum chaotic dynamics performed on a cloud-based IBM superconducting quantum processor, using the method described above to access operator-space measures of scrambling. The system in question, ibm_lagos [88], has seven qubits, arranged as illustrated in Fig. 3(a). In the main text we present results where five contiguous qubits are used to simulate a one-dimensional chaotic system using entangling gates arranged in a brickwork pattern [Fig. 3(b)]. Appendix D contains details of similar results that involve all seven qubits in the device, for which an alternative spacetime pattern of gates is needed.

A. Setup

The brickwork circuit is made up of entangling two-qubit gates, which we choose to be CNOT gates, combined with single-site unitaries. Each single-site gate is independently sampled from a uniform distribution over a discrete set of four gates $\{W_c : c = 1, \ldots, 4\}$. In terms of the native gates

of the quantum device ($\sqrt{X}$, X, and $R_\theta = e^{-i\theta Z/2}$), these are $W_1 = R_{\pi/4}\sqrt{X}R_{\pi/4}^\dagger$, $W_2 = R_{\pi/4}XR_{\pi/4}^\dagger$, $W_3 = \sqrt{X}R_{\pi/4}\sqrt{X}$, and $W_4 = \sqrt{X}R_{\pi/4}^\dagger\sqrt{X}$. In a given time step $t = 1, 2, \ldots$, CNOT gates are applied to pairs of qubits $(2j-1, 2j)$ for odd t (the first index is control and the second is target) and to $(2j, 2j+1)$ for even t. All qubits are then subjected to single-site unitaries $W_{c_{j,t}}$. This circuit is illustrated in Fig. 3(b). The indices $c_{j,t}$ have been sampled once for each j, t, and this configuration is used in all the data presented in this paper, i.e., we do not average over different single-qubit unitaries. This defines a time-dependent evolution channel $\mathcal{N}_{t=1,2,\ldots}$ that exhibits chaos.

In practice, for the quantum processor we use, running the same circuit many times is much faster than running many randomly generated circuits once each. For this reason, we alter the shadow protocol slightly: A random computational basis state $|\psi\rangle = \bigotimes_{j=1}^N |\hat{a}_j\rangle$ is used in place of the initial $|0^{\otimes N}\rangle$ (this can be done with a fixed circuit by preparing $|0^{\otimes N}\rangle$ and applying Hadamard gates to each qubit, followed by projective measurements of all qubits). The full circuit is sampled M_S times for a fixed choice of U_j and V_j, generating different $\hat{a}_j$ and $\hat{b}_j$ each time. The whole procedure is repeated for M_U different independently chosen bases. While the sampling errors in the final outcomes of observables are suboptimal for a fixed total number of runs $M_S M_U$ compared to the usual shadows protocol [42], we are able to reach a much higher total run count this way, thus achieving higher accuracy. We discuss the necessary alterations to the postprocessing methods and the influence on the scaling of errors in the Supplemental Material [85].

Other than the channel $\mathcal{N}_t$ itself, the full circuit involves single-qubit unitaries and measurements. To compensate for the imperfect measurement process, we run periodic calibration jobs, the data from which are used to apply measurement error mitigation techniques as described in, e.g., Ref. [89]. In principle, one could also employ a version of shadow tomography that counteracts the effects of errors in the unitaries U_j and V_j [90]; however, the single-qubit gate errors in ibm_lagos are on the order of 10^{-4}, so we assume that these unitaries are implemented perfectly.

The full shadow tomography protocol is executed on ibm_lagos with $M_S = 8192$ and $M_U = 900$ for t varying from 0 to 15. The values obtained from this data set are affected by both imperfections in $\mathcal{N}_t$ realized in the quantum device (noise) and the sampling error (i.e., the statistical fluctuations arising from the stochastic nature of shadow tomography). To help distinguish these two sources of error, we also generate another set of shadow data by running noise-free numerical simulations of the full circuits (Fig. 2) where all gates are perfectly accurate, and the measurement outcomes are sampled stochastically. This data set generates values that are affected by sampling error only. The same two sets of shadow data (which we label "Simulation" and "ibm_lagos") are used to calculate all the different physical quantities described in the following. We also compute the exact value of each quantity for noiseless $\mathcal{N}_t$, against which the shadow tomographic estimates will be compared. Throughout, we fix A and C to be individual qubits $A = \{1\}$ and $C = \{j_C\}$, where $j_C = 1, \ldots, N$.
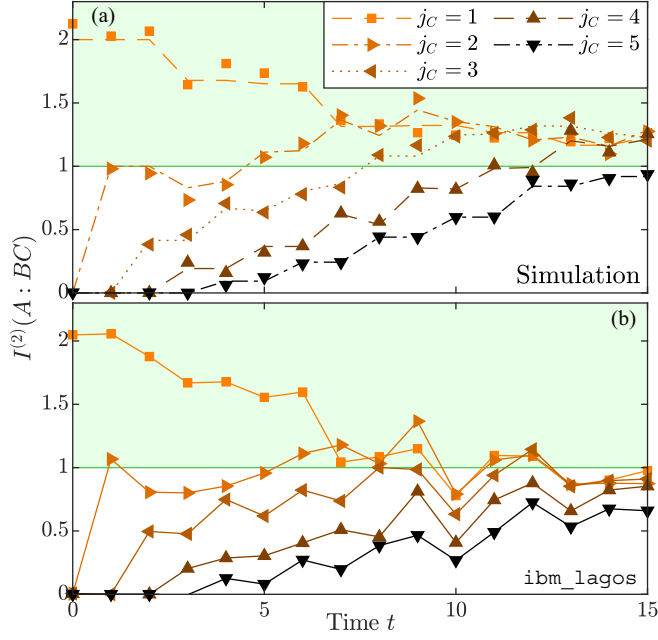
FIG. 4. Rényi mutual information [Eq. (2)], with $A = \{1\}$ and $C = \{j_C\}$. (a) Dashed lines indicate the exact value without noise or sampling error and points are estimations obtained using shadow postprocessing methods on data from numerical simulations of the full circuit (Fig. 2) without noise. The deviations between these two values can be used to estimate the typical size of the sampling errors that arise from the shadow tomography protocol. (b) Results obtained from ibm_lagos. Solid lines are to guide the eye. The region above the threshold $I^{(2)}(A : BC) > 1$ is shaded green (see Sec. II B for details).

B. Results

First, the Rényi mutual information $I^{(2)}(A : BC)$ is plotted in Fig. 4. At early times, the mutual information is large only for $j_C = 1$, reflecting the fact that the input A can only be reconstructed if one has access to the same qubit at the final time. At late times, the data from noiseless simulations saturate to comparable values for all choices of j_C, close to the value $I^* = 1.1945\ldots$ that would be expected if $\mathcal{N}_t$ were a global Haar random unitary (see Appendix A), thus confirming that information has scrambled. (For $j_C = 5$, this value is reached at a time just beyond the maximum t simulated on the quantum device.) The approach to this saturation value follows a light-cone structure: Qubits that are further away from A take a longer time to reach saturation. The results from the quantum processor agree well with simulations at early times. At later times we see an increasingly marked reduction of $I^{(2)}(A : BC)$ for all j_C. This is a consequence of the cumulative effects of noise in the execution of the time evolution $\mathcal{N}_t$, which reduces the fidelity of information transmission. For $j_C \leqslant 3$, we find values of $I^{(2)}(A : BC)$ above the threshold value of $|A| = 1$, which confirms that the quantum communication capacity of $\mathcal{N}_t^{A \to BC}$ is nonzero (see the preceding section). Even though the threshold is not exceeded for all qubits due to noise, the increase of $I^{(2)}(A : BC)$ confirms that information does indeed propagate to all qubits to some extent.
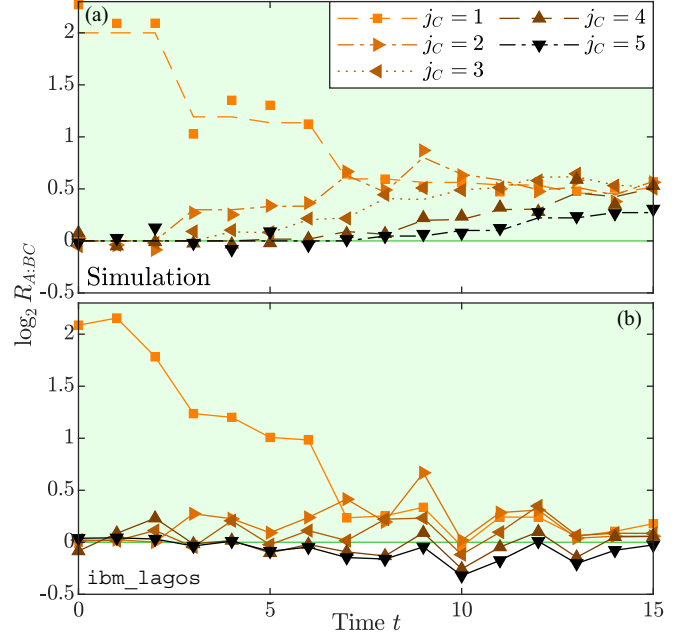

FIG. 5. Logarithm of the ratio $R_{A:BC} = p_{2,A:BC}^2 / p_{3,A:BC}$, where $A = \{1\}$ and $C = \{j_C\}$. Data are presented as in Fig. 4 for results from (a) simulation and (b) ibm_lagos. The region above the threshold $\log_2 R_{A:BC} > 0$ is shaded green (see Sec. II B for details).

The ratio of negativities $R_{A:BC}$ is plotted in Fig. 5. These show a similar pattern to the mutual information: The early-time values of $R_{A:BC}$ are large only for $j_C = 1$, and as time evolves the ratio tends towards saturation values that are comparable for all values of j_C, following a light-cone structure. From numerical simulations, we see that the threshold $R_{A:BC} > 1$ is achieved at earlier times than for the Rényi mutual information, suggesting that this criterion is more sensitive than the mutual information to the particular form of operator-space entanglement generated by the dynamics. On the other hand, the data from ibm_lagos show a more significant suppression of the signal, suggesting that the quantity in question may be more sensitive to noise.

Finally, in Fig. 6 we plot the cumulative sum $\sum_{l=0}^{k} D_l^C$ [Eq. (6)], which measures the proportion of the time-evolved operators $\sigma^\mu(t)$ that act nontrivially on at most k qubits, averaged over all nonidentity initial operators σ^μ with support in C. Here we fix $C = \{3\}$, the central qubit in the chain. Note that for unitary time evolution, the total operator weight $\sum_\nu |c^{\mu\nu}(t)|^2$ is conserved, which implies that $\sum_{l=0}^{N} D_l^C = 1$.

At early times, the operators have only evolved a small amount away from their single-qubit initial values and so the operator weight is dominated by the low-k sectors. As time evolves, an increasing amount of weight moves onto operators with more extended support. Eventually, once the system has fully scrambled, the evolved operators have weight roughly evenly distributed over the whole space of operators (excluding identity). The weights $D_k^C(t)$ are then well approximated by $(4^N - 1)^{-1} 3^k \binom{N}{k}$, which is the value that would be obtained from a uniform distribution over all $4^N - 1$ nontrivial operators. At these late times, the values obtained from ibm_lagos are again lower than the exact values due to noisy

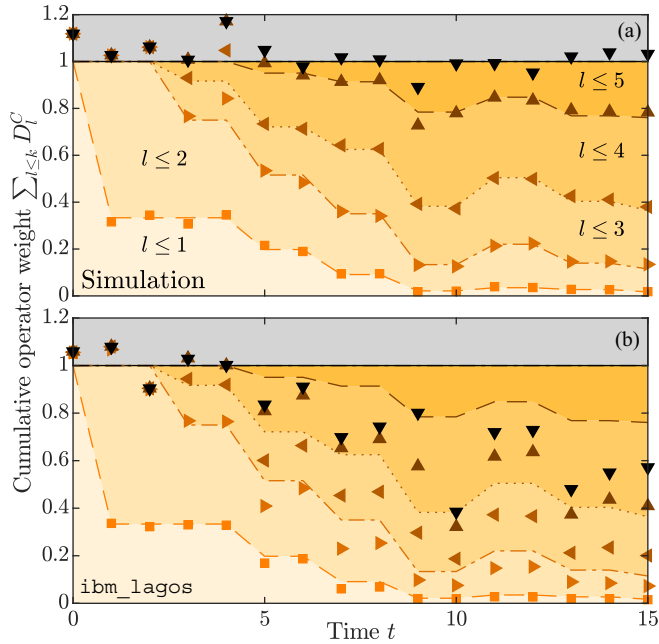FIG. 6. Evolution of the k-locality of time-evolved operators, as quantified by D_k^C [Eq. (6)]. Specifically, we plot the cumulative weight $\sum_{l \leqslant k} D_l^C$, which measures the total weight of the time-evolved operator acting nontrivially on at most k qubits, averaged over all nontrivial initial operators with support on C. We fix $C = \{3\}$, the central qubit in Fig. 3(a). The shaded areas and dashed lines indicate the exact values without sampling error or noise. Markers indicate shadow tomographic estimates calculated from the data sets obtained from (a) noise-free simulations and (b) the quantum device. The former are affected by sampling error only, while the latter are affected by both sampling error and noise.

nonunitary processes. Indeed, given that the dynamics of the quantum device is not perfectly unitary, the total operator weight $\sum_{\nu} |c^{\mu\nu}(t)|^2$ is expected to decrease with time, which is reflected in the data for $k = 5$.

V. DISCUSSION AND OUTLOOK

Using a combination of randomized state preparation and measurement, combined with the postprocessing techniques introduced in Ref. [42], we have evaluated various operator-space entanglement measures in a programmable quantum simulator. We constructed quantities that probe the fidelity of the Hayden-Preskill teleportation protocol [5], allowing us to unambiguously confirm that the system exhibits scrambling. Additionally, we used the same techniques to characterize operator growth, which can also be used to diagnose quantum chaos [57,58].

A related approach to diagnosing scrambling in experiments is to measure the decay of OTOCs [36–40,61]. However, present day quantum simulators are inevitably noisy, and dissipative effects can mimic this decay [41,91], as can mismatch between forward and backward time evolution. Thus, OTOC decay is at present not a truly verifiable diagnostic of scrambling to the same extent as many-body teleportation.

Compared to previous proposals to measure operator-space entanglement and teleportation fidelities [62,92], our method has the advantage that no additional ancilla qubits are needed. Not only does this reduce the hardware requirements in terms of system size, it also removes the need to control the dynamics of ancillas, which would otherwise need to be kept coherent and possibly time evolved in parallel [41]. Moreover, other than the time evolution $\mathcal{N}_t$ itself, the only additional gates required are single-qubit rotations, making the protocol particularly straightforward to implement on a wide variety of programmable quantum simulators. This simplicity is possible because our protocol does not require us to explicitly perform the decoding procedure for the many-body teleportation problem; rather, we can infer the existence of correlations between A and BC from statistical correlations between different measurement, which in turn informs us that teleportation is in principle possible.

In developing the protocol used here, we have focused on keeping experimental requirements to a minimum. However, other approaches that demand higher levels of experimental control may offer different advantages. In particular, one consequence of using randomized state preparation and measurement is the exponential scaling of the required number of repetitions M with the size of the region on which the Rényi entropy is evaluated; indeed, this sampling complexity is provably optimal with the given resources [42]. This is not an issue if one is interested in small regions within a large system, which is the situation for many studies of quantum thermalization, but may be problematic if one needs to consider large AC. Indeed, the ideal probes of many-body teleportation require access to an extensive number of inputs AB. Note, however, that one could consider correlations between A and $B'C$, where B' is a fixed size rather than the full complement of A, which will be good measures of early-time chaos; see also the modified OTOCs in Ref. [60].

One immediate generalization is to replace the random local unitaries U_j and V_j with global Clifford gates [42]. As argued in Ref. [55], the scaling of the required number of runs will be better, albeit still exponential. The larger number of gates required will make such a protocol more susceptible to decoherence and so noise-robust techniques would be required [90].

If the evolution in question $\mathcal{N}_t$ is known in advance, then further improvements to the scaling of M may be obtained using ancillary qubits. Roughly speaking, in these approaches the nonlocal correlations established during time evolution are distilled into smaller regions using some decoding procedure that requires knowledge of $\mathcal{N}_t$; these correlations can then be verified in a sample-efficient way. For instance, fast decoders for the Hayden-Preskill problem have been developed that use a doubled system [93]. Note that as the system size increases, so too will the complexity of these decoders, requiring increasingly high levels of coherence and gate fidelity. Thus, in current NISQ devices, there is a natural tradeoff between sample complexity and the necessary level of control over the system.

The quantities that one can directly access without using full tomography of $\mathcal{N}_t$ or an ansatz for $\rho^{AC}(t)$ [94] are limited to integer moments of the (doubled) density matrix $\rho_{op}(t)$. While the Rényi entropies $S^{(m)}(AC)$ and partially transposed

moments $p_{m,A:C}$ have less information-theoretic significance than, e.g., the von Neumann entropy, their experimental relevance makes it important to better understand their behavior in chaotic systems, which we leave to future work.

Recently, a protocol to measure the spectral form factor, a quantity that can be used to diagnose chaos in time-periodic systems [95], has been proposed, which also uses randomized state preparation and measurement [96]. There the initial and final unitaries appearing in Fig. 2 are related via $U_j = V_j^\dagger$. It would be interesting to consider other ways of introducing correlations between different random unitaries in such protocols, which could give access to different properties of the time-evolution channel.

Operator-space entanglement also plays an important role in contexts beyond quantum chaos. For instance, the mutual information between initial and final states can be used as a probe of entanglement phase transitions in monitored quantum circuits [97–100]. Analogous quantities can also be used to detect quantized chiral information propagation at the edge of anomalous Floquet topological phases [101–104]. The protocol we employ here could therefore be used as a means to verify experimental realizations of these phenomena.

Note added. Recently, Refs. [82,83] appeared, where similar proposals to generalize shadow tomography to channels were given.

In compliance with EPSRC policy framework on research data, data obtained from numerical simulations and experiments on ibm_lagos will be made publicly accessible via Zenodo [105].

APPENDIX A: PROPERTIES OF THE RéNYI MUTUAL INFORMATION (2)

In this Appendix we prove the claims made in the main text regarding properties of Rényi mutual information [Eq. (2)] when the underlying state is an operator state $\rho_{\mathrm{op}}(t)$, including our claim that the quantum capacity of a channel must be nonzero when the corresponding operator-space Rényi mutual information exceeds its maximum classical value. We will refer explicitly to the quantity $I^{(m)}(A:C) := S^{(m)}(A) + S^{(m)}(C) - S^{(m)}(AC)$, where A is a subset of inputs and C is a subset of outputs [as in Fig. 1(b)]; however, our claims continue to hold if A and C are replaced by subsets that contain combinations of inputs and outputs, provided the reduced density matrix on at least one of the subsets is maximally mixed. For example, in the main text we consider $I^{(m)}(A:BC)$, which falls under this category since the reduced density matrix $\rho^A = \mathbb{I}_A/2^{|A|}$ is maximally mixed. For the purposes of this Appendix, we leave all t dependence implicit. We

denote the Hilbert space dimensions of A and C by d_A and d_C, respectively.

While the definition of the Rényi mutual information that we use here [Eq. (2)] generalizes the von Neumann mutual information in a natural way, it is not always a good measure of the correlations present in a given state. For example, in certain cases it can even be negative [79,80]. [Because of this, other related quantities have been proposed that are sometimes referred to as Rényi mutual information [78]; here we will use this term exclusively for the quantity (2).] However, when the reduced density matrix for either A or C is maximally mixed, as occurs in the cases under consideration, it was noted that $I^{(m)}(A:C)$ is non-negative [46]. We argue that this can be made stronger.

Theorem. For any density operator ρ^{AC} satisfying $\rho^A := \mathrm{Tr}_C \rho^{AC} = \mathbb{I}_A/d_A$ or $\rho^C = \mathbb{I}_C/d_C$, the Rényi mutual information satisfies

$$I^{(m)}(A:C) \geqslant 0 \,\forall\, m = 2, 3, \ldots \tag{A1}$$

with equality if and only if the density operator factorizes as $\rho^{AC} = \rho^A \otimes \rho^C$.

This theorem establishes $I^{(m)}(A:C)$ as a sensible measure of how much ρ^{AC} fails to factorize and hence the degree to which A and C are correlated. We only explicitly consider integer $m \geqslant 2$ here, since these are the quantities that can be measured experimentally.

We assume that ρ^A is maximally mixed; the alternative case where ρ^C is maximally mixed then follows from the symmetry of $I^{(m)}(A:C)$. Our proof relies on the observation

$$\int_{\mathrm{Haar}} dU_1 \cdots dU_m \mathrm{Tr}[(U_1 \otimes \mathbb{I}_C)\rho_{AC}(U_1 \otimes \mathbb{I}_C)^\dagger \times \cdots$$
$$\times (U_m \otimes \mathbb{I}_C)\rho_{AC}(U_m \otimes \mathbb{I}_C)^\dagger] = \mathrm{Tr}\left[\left(\frac{\mathbb{I}_A}{d_A}\right)^m \otimes (\rho_C)^m\right], \tag{A2}$$

where the integration variables are unitary matrices $\{U_i \in U(d_A)\}$ acting on A and the integrals are taken over the Haar measure. The above is a consequence of the standard identity $\int_{\mathrm{Haar}} dU\, U\mathcal{O}U^\dagger = (\mathrm{Tr}[\mathcal{O}]/d)\mathbb{I}_d$ for $d \times d$ matrices $\mathcal{O}$ [106]. We seek to prove $\mathrm{Tr}[(\rho^{AC})^m] \geqslant \mathrm{Tr}[(\rho^A)^m \otimes (\rho^C)^m]$, which will in turn imply (A1). Since the integration measure over each U_i is normalized $\int_{\mathrm{Haar}} dU_i = 1$ and $\rho^A = \mathbb{I}_A/d_A$, we have

$$\mathrm{Tr}[(\rho^{AC})^m] - \mathrm{Tr}[(\rho^A)^m \otimes (\rho^C)^m]$$
$$= \int_{\mathrm{Haar}} dU_1 \cdots dU_m \{\mathrm{Tr}[(\rho^{AC})^m]$$
$$- \mathrm{Tr}[U_1 \rho^{AC} U_1^\dagger \cdots U_m \rho^{AC} U_m^\dagger]\}, \tag{A3}$$

where we leave the factors of $\mathbb{I}_C$ implicit. The integrand on the right-hand side is non-negative by the following lemma.

Lemma. If ρ is a complex Hermitian positive-semidefinite matrix and $\{U_i\}$ are unitary matrices of the same size, then

$$|\mathrm{Tr}[U_1 \rho U_1^\dagger \cdots U_m \rho U_m^\dagger]| \leqslant \mathrm{Tr}[\rho^m] \tag{A4}$$

with equality if and only if $U_1 \rho U_1^\dagger = \cdots = U_m \rho U_m^\dagger$.

Proof. We first note that $|\mathrm{Tr}[A]| \leqslant \mathrm{Tr}[|A|]$ for all square matrices A, where $|A| := (A^\dagger A)^{1/2}$, with equality if and only if A is Hermitian positive semidefinite. Setting $A = A_1 \cdots A_m$

where $A_j = U_j \rho U_j^\dagger$, we then use a generalization of Hölder's inequality proved in Ref. [107],

$$\text{Tr}[|A_1 \cdots A_m|] \leqslant \prod_{a=1}^{m} \{\text{Tr}[(A_a)^{p_a}]\}^{1/p_a} \qquad \text{(A5)}$$

for any positive real numbers p_a satisfying $\sum_a (1/p_a) = 1$, with equality if and only if $A_1 = \cdots = A_m$. Equation (A4) then follows by setting $p_a = m$ for $a = 1, \ldots, m$ so that $\text{Tr}[(A_a)^{p_a}] = \text{Tr}[\rho^m]$. ∎

This completes our proof that $I^{(m)}(A : C)$ is non-negative for the states under consideration. The fact that $I^{(m)}(A : C)$ vanishes for factorizable ρ^{AC} follows immediately from its definition. Conversely, if $I^{(m)}(A : C) = 0$, then the integrand in (A3) must vanish everywhere, which implies that $(U \otimes \mathbb{I}_C)\rho^{AC}(U \otimes \mathbb{I}_C)^\dagger = \rho^{AC}$ for all $U \in U(d_A)$. This can only be true if $\rho^{AC} \propto \mathbb{I}_A \otimes \rho^C$, which completes our proof.

Having established the above theorem, we now provide the proof of the claims we made in Sec. II B regarding the threshold values for $I^{(m)}(A : BC)$ and $R_{A:BC}$. In its most general form, we have the following claim.

Claim. The quantum capacity of a channel $\mathcal{N}^{A \to B}$ is nonzero if the operator-space Rényi mutual information satisfies $I^{(m)}(A : B) > |A|$. If the input Hilbert space dimension $|\mathcal{H}_A| = 2$, then the same conclusion can be made whenever the ratio of partially transposed moments $R_{A:B} := p_{2,A:B}^2/p_{3,A:B}$ exceeds unity.

The statements made in the main text then follow from applying the above to $\mathcal{N}_t^{A \to B'C}$.

Proof. First, we consider the case where $I^{(m)}(A : B) > |A|$. Here we will rely somewhat on the notion of majorization; see, e.g., Ref. [108] for a full introduction. An $n_X \times n_X$ Hermitian matrix X majorizes an $n_Y \times n_Y$ Hermitian matrix Y if their traces are equal and the sum of the kth largest eigenvalues of X is greater than or equal to the sum of the kth largest eigenvalues of Y for $k = 1, 2, \ldots, \min(n_X, n_Y)$. This relation is denoted by $X \succeq Y$. A function f from matrices to real numbers is called Schur convex if and only if $X \succeq Y \Rightarrow f(X) \geqslant f(Y)$.

Since A is maximally mixed, our starting point $I^{(m)}(A : B) > |A|$ is equivalent to $\text{Tr}[(\rho^{AB})^m] > \text{Tr}[(\rho^B)^m]$, where ρ^{AB} is the operator state for the channel $\mathcal{N}^{A \to B}$ [see Eq. (1)] and $\rho^B = \text{Tr}_A \rho^{AB}$. It is straightforward to show that the map $\rho \mapsto \text{Tr}[\rho^m]$ is Schur convex, which implies that $\rho^{AB} \npreceq \rho^B$. In Ref. [109] it was shown that separable states satisfy $\rho^{AB} \preceq \rho^B$, and so the operator state must be bipartite entangled whenever $I^{(m)}(A : B) > |A|$. Moreover, in Ref. [110] a stronger result was proved: Violation of the separability criterion $\rho^{AB} \preceq \rho^B$ implies violation of the so-called reduction criterion [111].

States which violate the reduction criterion must possess *distillable* entanglement, meaning that many copies of the state can be converted into a smaller number of pure EPR pairs using local operations and classical communication [112].

The above implies that if the operator state ρ^{AB} satisfies $I^{(m)}(A : B) > |A|$, then pure EPR pairs can be distilled from many copies of ρ^{AB} (each of which can be prepared from a single use of the channel $\mathcal{N}^{A \to B}$) using the protocol described in Ref. [111], which requires a one-way classical communication channel from sender A to receiver B. The ability to generate EPR pairs from multiple uses of a channel assisted by one-way classical communication is equivalent to being able to reliably transmit the same number of qubits from A to B using the same resources [113]. Since the quantum channel capacity assisted by one-way classical communication is equal to the unassisted capacity [113,114], we conclude that the quantum capacity of any channel $\mathcal{N}^{A \to B}$ must be nonzero whenever the operator state ρ^{AB} satisfies $I^{(m)}(A : B) > |A|$.

For the ratio of partially transposed moments $R_{A:B}$ [Eq. (3)], our argument follows a similar line. In Ref. [77] it was shown that if a bipartite state ρ^{AB} satisfies $R_{A:B} > 1$, then the Peres criterion [115] must be violated, which is a sufficient but not necessary condition for the existence of bipartite entanglement in ρ^{AB}. Given that the Hilbert space dimension $|\mathcal{H}_A| = 2$, violation of the Peres criterion implies that the entanglement in ρ^{AB} is distillable [116]. Again using the equivalence between generation of pure EPR pairs and transmission of quantum states, we conclude that the quantum capacity of $\mathcal{N}^{A \to B}$ must be nonzero.

Finally, it is helpful to evaluate $I^{(m)}(A : C)$ for the case where the time evolution is a global Haar random unitary, which is maximally chaotic. A simple estimate for the average $\langle I^{(m)}(A : C) \rangle_{U_t}$ (angular brackets denote the expectation value over all unitary evolutions U_t with respect to the Haar measure) can be obtained by approximating $\langle \log_2 \text{Tr}[\rho^{AC}(t)^m] \rangle_{U_t} \approx \log_2 \langle \text{Tr}[\rho^{AC}(t)^m] \rangle_{U_t}$, the right-hand side of which can be evaluated using standard expressions for integrals over the Haar measure [106]. This assumes that fluctuations of $\text{Tr}[\rho^{AC}(t)^m]$ between different Haar random unitaries are small. For the simplest case of $m = 2$, for a system of N q-level systems ($q = 2$ for our case of qubits), we find

$$\langle \text{Tr}[\rho^{AC}(t)^2] \rangle_{U_t} = \frac{1}{q^N(q^{2N} - 1)}[q^N(q^{|BD|} + q^{|AC|}) - (q^{|AD|} + q^{|BC|})]. \qquad \text{(A6)}$$

This can be used to estimate the mean value of $I^{(2)}(A : BC)$, which we argue in the main text probes the fidelity of the Hayden-Preskill teleportation protocol

$$\langle I^{(m)}(A : BC) \rangle_{U_t} \approx |AC| \log_2 q - \log_2 \left(\frac{q^{2N}(q^{|A|-|C|} + q^{|C|-|A|} - q^{-|AC|}) - q^{|AC|}}{q^{2N} - 1} \right). \qquad \text{(A7)}$$

In the case of interest $|A| = |C|$, this becomes

$$\langle I^{(m)}(A : BC) \rangle_{U_t} \cong |AC| \log_2 q - \log_2 \left(\frac{q^{2N}(2 - q^{-|AC|}) - q^{|AC|}}{q^{2N} - 1} \right). \qquad \text{(A8)}$$

The first term on the right-hand side is the maximum value for the Rényi mutual information. The second term, describing deviations from the maximum value, remains order one when one takes $|N| \to \infty$ while keeping $|A| = |C|$ fixed. This is consistent with the expectation that information about the initial state of A can be recovered even if one only has access to a vanishing fraction of outputs C (this corresponds to the amount of Hawking radiation in the Hayden-Preskill protocol [5]). Evaluating (A8) for the case $N = 5$, $q = 2$, and $|A| = |C| = 1$ (the parameters used for the data plotted in Fig. 4), we find $I^{(2)}(A : BC) \approx 1.1945\ldots$.

APPENDIX B: PROOF OF EQ. (6)

Here we prove the relationship between the quantities $D_k^C(t)$, which measure the k-locality of time-evolved operators that initially have support in C, and the operator purities $\text{Tr}[\rho^{AC}(t)^2]$. First, trace preservation implies that $\mathcal{N}^\dagger[\mathbb{I}] = \mathbb{I}$, which in turn gives $c^{Iv}(t) = \delta_{v,I}$, where I labels the identity Pauli string. Thus, for $k \geqslant 1$, the restriction $\mu \neq I$ in the sum in the first line of (6) can be removed. Then we consider the sum of operator purities over all subsets of qubits A of fixed size $|A| = r$,

$$E_r^C(t) := \frac{2^{|C|+r}}{2^{|C|}-1} \sum_{A \subseteq S; |A|=r} \text{Tr}[\rho^{AC}(t)^2] \quad \text{(B1)}$$

$$= \frac{1}{2^{|C|}-1} \sum_{k=0}^{r} \binom{N-k}{r-k} \sum_{\mu \in C} \sum_{\nu:|\sigma^\nu|=k} |c^{\mu\nu}(t)|^2 \quad \text{(B2)}$$

$$= \sum_{k=0}^{r} \binom{N-k}{N-r} D_k^C, \quad \text{(B3)}$$

where for convenience we alter the definition of D_k^C for $k = 0$ to be $D_0^C = \text{Tr}[\rho^C(t)^2]/(2^{|C|}-1)$, which differs from the expression (6) in the inclusion of the term $\mu = I$. The above follows from counting the number of subregions A that support a Pauli string that acts nontrivially on k qubits. This establishes a linear relationship between the sums $E_r^C(t)$ and the quantities of interest D_k^C, which can be inverted. The inverse of the lower triangular matrix $[L]_{rk} = \binom{N-k}{N-r}$ $(r \geqslant k)$ is simply given by $[L^{-1}]_{kr} = (-1)^{k+r}\binom{N-r}{N-k}$ $(k \geqslant r)$; this can be proved using the relation $\sum_{m=i}^{j}(-1)^{j+m}\binom{j}{m}\binom{m}{i} = \delta_{ij}$. This gives $D_k^C = \sum_{r=0}^{k}(-1)^{r+k}\binom{N-r}{N-k}E_r^C(t)$, which can be easily manipulated to give Eq. (6).

APPENDIX C: JUSTIFICATION OF EQ. (7)

In this Appendix we prove that the quantity (7) is indeed an unbiased estimator of the operator state $\rho_{\text{op}}(t)$, i.e., $\mathbb{E}[\hat\rho_{\text{op}}(t)] = \rho_{\text{op}}(t)$, where the expectation value is taken over the joint distribution of unitaries U_j and V_j and outcomes $\hat{b}_j$. This can be done relatively straightforwardly using the graphical equation shown in Fig. 7. First, suppose that one could explicitly construct $\rho_{\text{op}}(t)$ in the experiment; then one could perform conventional shadow tomography, where unitaries U and V are applied to Q_{in} and Q_{out}, respectively, with outcomes $\hat{a}, \hat{b} \in \{0,1\}^{\times N}$. This is shown in Fig. 7(a). Using the property of the maximally mixed state $(\mathcal{O}_{\text{in}}^T \otimes \mathbb{I}_{\text{out}})|\Phi\rangle =$
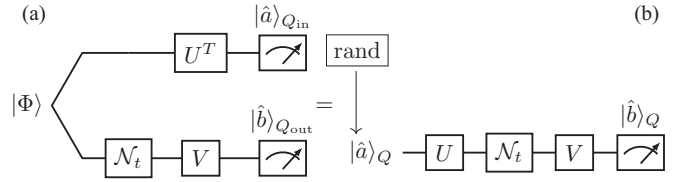

FIG. 7. (a) Conventional shadow tomography on the operator state $\rho_{\text{op}}(t) = (\text{id}_{\text{in}} \otimes \mathcal{N}_t)[\Phi]$. The distribution of unitaries U and V and measurement outcomes $\hat{a}$ and $\hat{b}$ are the same as that of (b) a hybrid classical-quantum process, where $\hat{a}$ are sampled from a uniform distribution and then used as the input for a quantum circuit.

$(\mathbb{I}_{\text{in}} \otimes \mathcal{O}_{\text{out}})|\Phi\rangle$, one can push the unitary $U = \bigotimes_j U_j$ acting on Q_{in} onto the other half of the doubled system. This makes it clear that the distribution of measurements $\{\hat{a}\}$ on the input qubits is uniform over $\{0,1\}^{\times N}$. Thus, we can sample $\hat{a}$ using a classical computer and use it as the input to a circuit that only requires a single copy of the system [Fig. 7(b)]. The joint distribution of $\hat{a}$, $\hat{b}$, U, and V will be exactly the same as that of state shadow tomography on $\rho_{\text{op}}(t)$, which allows us to construct an unbiased estimator of $\rho_{\text{op}}(t)$ in the usual way [42].

Finally, we note that the variables $\hat{a}$ and U only appear in the combination $U|\hat{a}\rangle$ in both the circuit and the shadow tomography estimator of the density matrix. Thus, we need only ensure that the ensemble of inputs to the channel $\mathcal{N}_t$ has the correct distribution. In our case U is distributed uniformly over products of single-qubit Clifford operations; we can therefore replace $U|\hat{a}\rangle$ with $U|0^{\otimes N}\rangle$ without modifying the appropriate distribution. This justifies the form of Eq. (7).

APPENDIX D: RESULTS FOR $N = 7$ QUBITS

In this Appendix we describe a circuit model of dynamics that uses all seven qubits of the quantum device ibm_lagos and present results obtained from the shadow protocol.

To generate chaotic dynamics, we use a circuit design made up of the same gates as the setup presented in the main text [Fig. 3(b)], namely, CNOT gates and single-qubit gates independently sampled from the discrete set $\{W_c : c = 1, \ldots, 4\}$. As before, each time step is made up of a layer of single-qubit unitaries acting on all qubits followed by a layer of CNOT gates. The arrangement of CNOT gates changes each time step, repeating itself after a period of three steps, as illustrated in Fig. 8. This ensures that entanglement can be generated between any two qubits after a sufficient amount of time.

After running the shadow tomography protocol with the same parameters as before ($N_U = 900$ and $N_M = 8192$), the Rényi mutual information $I^{(2)}(A : BC)$ is computed, where we set $A = \{1\}$, the top left qubit in Fig. 8. We also generate a set of shadow data by simulating the full circuit without noise on a classical computer, for comparison. The results are presented in Fig. 9. Initially, correlations are only present for $j_C = 1$, whereas at later times these correlations are distributed across the entire system, thus confirming that information has been scrambled. As before, the values from
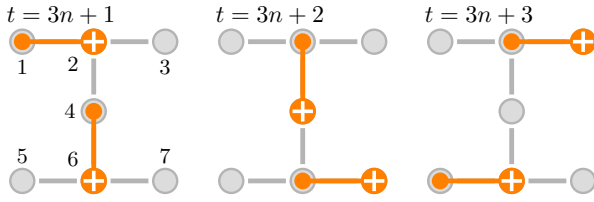
FIG. 8. Configuration of the CNOT layers in the model of chaotic dynamics that uses all seven qubits of ibm_lagos. The pattern of CNOT gates repeats every three time steps. Small orange circles denote the control qubits and large orange circles with a plus are the target qubits. Note that the numbering of the qubits (indicated in the leftmost panel) differs from that used in the text.
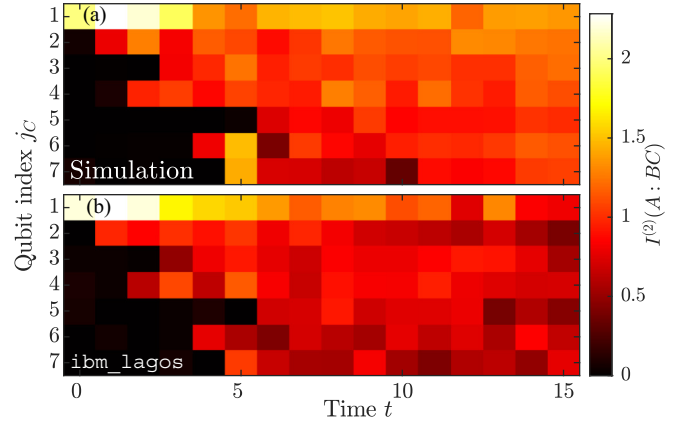

FIG. 9. Color plot of the second Rényi mutual information $I^{(2)}(A:BC)$ for the circuit model of dynamics described in Appendix D, using all seven qubits of ibm_lagos. We fix $A = \{1\}$ [the bottom right qubit in Fig. 3(a)] and $C = \{j_C\}$, where j_C is varied. (a) Results obtained from noiseless classical simulations of the time evolution $\mathcal{N}_t$ and the shadow protocol. By averaging the deviation of these data from the exact value of $I^{(2)}(A:BC)$ for the circuit in question, we obtain an estimate of the statistical fluctuations due to the shadow protocol of 0.07. (b) Data obtained from ibm_lagos.

ibm_lagos at later times are systematically below those from classical simulations, due to noisy processes that disturb the propagation of information.

The region ABC involves more qubits than that used for the $N = 5$ setup described in the main text and so we expect to incur larger statistical errors when computing the operator-space Rényi entropies and in turn $I^{(2)}(A:BC)$. The size of these errors can be estimated by looking at the deviation of the values from noiseless classical simulations of the shadow protocol, compared with the exact values of the mutual information. Averaging across all times t and choices of j_C, we find a mean relative error in the value of $\mathrm{Tr}[\rho_{ABC}^2]$ of 0.05 and an absolute error in $I^{(2)}(A:BC)$ of 0.07. Evidently, even for

regions as large as $|ABC| = 8$, it is possible to estimate Rényi entropies and quantities derived thereof to a good accuracy using a reasonable number of shots.

[1] J. M. Deutsch, Phys. Rev. A **43**, 2046 (1991).

[2] M. Srednicki, Phys. Rev. E **50**, 888 (1994).

[3] H. Tasaki, Phys. Rev. Lett. **80**, 1373 (1998).

[4] M. Rigol, V. Dunjko, and M. Olshanii, Nature (London) **452**, 854 (2008).

[5] P. Hayden and J. Preskill, J. High Energy Phys. 09 (2007) 120.

[6] Y. Sekino and L. Susskind, J. High Energy Phys. 10 (2008) 065.

[7] S. H. Shenker and D. Stanford, J. High Energy Phys. 03 (2014) 067.

[8] S. W. Hawking, Phys. Rev. D **14**, 2460 (1976).

[9] J. Benhelm, G. Kirchmair, C. F. Roos, and R. Blatt, Nat. Phys. **4**, 463 (2008).

[10] D. Nigg, M. Müller, E. A. Martinez, P. Schindler, M. Hennrich, T. Monz, M. A. Martin-Delgado, and R. Blatt, Science **345**, 302 (2014).

[11] J. Zhang, G. Pagano, P. W. Hess, A. Kyprianidis, P. Becker, H. Kaplan, A. V. Gorshkov, Z.-X. Gong, and C. Monroe, Nature (London) **551**, 601 (2017).

[12] N. Friis, O. Marty, C. Maier, C. Hempel, M. Holzäpfel, P. Jurcevic, M. B. Plenio, M. Huber, C. Roos, R. Blatt, and B. Lanyon, Phys. Rev. X **8**, 021012 (2018).

[13] R. Barends, J. Kelly, A. Megrant, A. Veitia, D. Sank, E. Jeffrey, T. C. White, J. Mutus, A. G. Fowler, B. Campbell *et al.*, Nature (London) **508**, 500 (2014).

[14] J. Kelly, R. Barends, A. G. Fowler, A. Megrant, E. Jeffrey, T. C. White, D. Sank, J. Y. Mutus, B. Campbell, Y. Chen *et al.*, Nature (London) **519**, 66 (2015).

[15] N. Ofek, A. Petrenko, R. Heeres, P. Reinhold, Z. Leghtas, B. Vlastakis, Y. Liu, L. Frunzio, S. Girvin, L. Jiang *et al.*, Nature (London) **536**, 441 (2016).

[16] G. Wendin, Rep. Prog. Phys. **80**, 106001 (2017).

[17] X. Mi, M. Ippoliti, C. Quintana, A. Greene, Z. Chen, J. Gross, F. Arute, K. Arya, J. Atalaya, R. Babbush *et al.*, Nature (London) **601**, 531 (2022).

[18] H. Weimer, M. Müller, I. Lesanovsky, P. Zoller, and H. P. Büchler, Nat. Phys. **6**, 382 (2010).

[19] J. T. Barreiro, M. Müller, P. Schindler, D. Nigg, T. Monz, M. Chwalla, M. Hennrich, C. F. Roos, P. Zoller, and R. Blatt, Nature (London) **470**, 486 (2011).

[20] D. Barredo, S. de Léséleuc, V. Lienhard, T. Lahaye, and A. Browaeys, Science **354**, 1021 (2016).

[21] M. Endres, H. Bernien, A. Keesling, H. Levine, E. R. Anschuetz, A. Krajenbrink, C. Senko, V. Vuletic, M. Greiner, and M. D. Lukin, Science **354**, 1024 (2016).

[22] H. Bernien, S. Schwartz, A. Keesling, H. Levine, A. Omran, H. Pichler, S. Choi, A. S. Zibrov, M. Endres, M. Greiner *et al.*, Nature (London) **551**, 579 (2017).

[23] S. Ebadi, T. T. Wang, H. Levine, A. Keesling, G. Semeghini, A. Omran, D. Bluvstein, R. Samajdar, H. Pichler, W. W. Ho *et al.*, Nature (London) **595**, 227 (2021).

[24] J. Preskill, Quantum **2**, 79 (2018).

[25] F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends, R. Biswas, S. Boixo, F. G. Brandao, D. A. Buell *et al.*, Nature (London) **574**, 505 (2019).

[26] H.-S. Zhong, H. Wang, Y.-H. Deng, M.-C. Chen, L.-C. Peng, Y.-H. Luo, J. Qin, D. Wu, X. Ding, Y. Hu, P. Hu, X.-Y. Yang, W.-J. Zhang, H. Li, Y. Li, X. Jiang, L. Gan, G. Yang, L. You, Z. Wang *et al.*, Science **370**, 1460 (2020).

[27] A. K. Ekert, C. M. Alves, D. K. L. Oi, M. Horodecki, P. Horodecki, and L. C. Kwek, Phys. Rev. Lett. **88**, 217901 (2002).

[28] C. M. Alves and D. Jaksch, Phys. Rev. Lett. **93**, 110501 (2004).

[29] A. J. Daley, H. Pichler, J. Schachenmayer, and P. Zoller, Phys. Rev. Lett. **109**, 020505 (2012).

[30] H. Pichler, L. Bonnes, A. J. Daley, A. M. Läuchli, and P. Zoller, New J. Phys. **15**, 063003 (2013).

[31] R. Islam, R. Ma, P. M. Preiss, M. E. Tai, A. Lukin, M. Rispoli, and M. Greiner, Nature (London) **528**, 77 (2015).

[32] A. I. Larkin and Y. N. Ovchinnikov, Sov. Phys. JETP **28**, 1200 (1969).

[33] A. Kitaev, A simple model of quantum holography (part 1), http://online.kitp.ucsb.edu/online/entangled15/kitaev/.

[34] D. A. Roberts, D. Stanford, and L. Susskind, J. High Energy Phys. 03 (2015) 051.

[35] I. L. Aleiner, L. Faoro, and L. B. Ioffe, Ann. Phys. (NY) **375**, 378 (2016).

[36] J. Li, R. Fan, H. Wang, B. Ye, B. Zeng, H. Zhai, X. Peng, and J. Du, Phys. Rev. X **7**, 031011 (2017).

[37] M. Gärttner, J. G. Bohnet, A. Safavi-Naini, M. L. Wall, J. J. Bollinger, and A. M. Rey, Nat. Phys. **13**, 781 (2017).

[38] K. X. Wei, C. Ramanathan, and P. Cappellaro, Phys. Rev. Lett. **120**, 070501 (2018).

[39] M. K. Joshi, A. Elben, B. Vermersch, T. Brydges, C. Maier, P. Zoller, R. Blatt, and C. F. Roos, Phys. Rev. Lett. **124**, 240505 (2020).

[40] X. Mi *et al.* (Google Quantum AI), Science **374**, 1479 (2021).

[41] B. Yoshida and N. Y. Yao, Phys. Rev. X **9**, 011006 (2019).

[42] H.-Y. Huang, R. Kueng, and J. Preskill, Nat. Phys. **16**, 1050 (2020).

[43] P. Zanardi, C. Zalka, and L. Faoro, Phys. Rev. A **62**, 030301(R) (2000).

[44] P. Zanardi, Phys. Rev. A **63**, 040304(R) (2001).

[45] P. Hosur, X.-L. Qi, D. A. Roberts, and B. Yoshida, J. High Energy Phys. 02 (2016) 004.

[46] Y. D. Lensky and X.-L. Qi, J. High Energy Phys. 06 (2019) 025.

[47] T. Zhou and D. J. Luitz, Phys. Rev. B **95**, 094206 (2017).

[48] J. Dubail, J. Phys. A: Math. Theor. **50**, 234001 (2017).

[49] E. Iyoda and T. Sagawa, Phys. Rev. A **97**, 042330 (2018).

[50] R. Pal and A. Lakshminarayan, Phys. Rev. B **98**, 174304 (2018).

[51] L. Nie, M. Nozaki, S. Ryu, and M. T. Tan, J. Stat. Mech. (2019) 093107.

[52] O. Schnaack, N. Bölter, S. Paeckel, S. R. Manmana, S. Kehrein, and M. Schmitt, Phys. Rev. B **100**, 224302 (2019).

[53] B. Bertini and L. Piroli, Phys. Rev. B **102**, 064305 (2020).

[54] G. Styliaris, N. Anand, and P. Zanardi, Phys. Rev. Lett. **126**, 030601 (2021).

[55] A. Elben, B. Vermersch, M. Dalmonte, J. I. Cirac, and P. Zoller, Phys. Rev. Lett. **120**, 050406 (2018).

[56] T. Brydges, A. Elben, P. Jurcevic, B. Vermersch, C. Maier, B. P. Lanyon, P. Zoller, R. Blatt, and C. F. Roos, Science **364**, 260 (2019).

[57] A. Nahum, S. Vijay, and J. Haah, Phys. Rev. X **8**, 021014 (2018).

[58] C. W. von Keyserlingk, T. Rakovszky, F. Pollmann, and S. L. Sondhi, Phys. Rev. X **8**, 021013 (2018).

[59] V. Khemani, A. Vishwanath, and D. A. Huse, Phys. Rev. X **8**, 031057 (2018).

[60] B. Vermersch, A. Elben, L. M. Sieberer, N. Y. Yao, and P. Zoller, Phys. Rev. X **9**, 021061 (2019).

[61] R. J. Garcia, Y. Zhou, and A. Jaffe, Phys. Rev. Research **3**, 033155 (2021).

[62] K. A. Landsman, C. Figgatt, T. Schuster, N. M. Linke, B. Yoshida, N. Y. Yao, and C. Monroe, Nature (London) **567**, 61 (2019).

[63] M. S. Blok, V. V. Ramasesh, T. Schuster, K. O'Brien, J. M. Kreikebaum, D. Dahlen, A. Morvan, B. Yoshida, N. Y. Yao, and I. Siddiqi, Phys. Rev. X **11**, 021010 (2021).

[64] A. Peruzzo, J. McClean, P. Shadbolt, M.-H. Yung, X.-Q. Zhou, P. J. Love, A. Aspuru-Guzik, and J. L. O'brien, Nat. Commun. **5**, 4213 (2014).

[65] J. Carolan, C. Harrold, C. Sparrow, E. Martín-López, N. J. Russell, J. W. Silverstone, P. J. Shadbolt, N. Matsuda, M. Oguma, M. Itoh, G. D. Marshall, M. G. Thompson, J. C. F. Matthews, T. Hashimoto, J. L. O'Brien, and A. Laing, Science **349**, 711 (2015).

[66] F. Flamini, N. Spagnolo, and F. Sciarrino, Rep. Prog. Phys. **82**, 016001 (2019).

[67] A. Jamiołkowski, Rep. Math. Phys. **3**, 275 (1972).

[68] M.-D. Choi, Linear Algebra Appl. **10**, 285 (1975).

[69] B. Schumacher and M. D. Westmoreland, Phys. Rev. A **56**, 131 (1997).

[70] A. Holevo, IEEE Trans. Inf. Theory **44**, 269 (1998).

[71] G. Vidal and R. F. Werner, Phys. Rev. A **65**, 032314 (2002).

[72] M. B. Plenio, Phys. Rev. Lett. **95**, 090503 (2005).

[73] J. Kudler-Flam, M. Nozaki, S. Ryu, and M. T. Tan, J. High Energy Phys. 01 (2020) 031.

[74] M. A. Nielsen and I. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, 2010).

[75] R. Pisarczyk, Z. Zhao, Y. Ouyang, V. Vedral, and J. F. Fitzsimons, Phys. Rev. Lett. **123**, 150502 (2019).

[76] D. Gross, Y.-K. Liu, S. T. Flammia, S. Becker, and J. Eisert, Phys. Rev. Lett. **105**, 150401 (2010).

[77] A. Elben, R. Kueng, H.-Y. R. Huang, R. van Bijnen, C. Kokail, M. Dalmonte, P. Calabrese, B. Kraus, J. Preskill, P. Zoller, and B. Vermersch, Phys. Rev. Lett. **125**, 200501 (2020).

[78] M. M. Wilde, A. Winter, and D. Yang, Commun. Math. Phys. **331**, 593 (2014).

[79] M. Berta, K. P. Seshadreesan, and M. M. Wilde, Phys. Rev. A **91**, 022333 (2015).

[80] S. O. Scalet, Á. M. Alhambra, G. Styliaris, and J. I. Cirac, Quantum **5**, 541 (2021).

[81] H.-Y. Hu, S. Choi, and Y.-Z. You, arXiv:2107.04817.

[82] R. Levy, D. Luo, and B. K. Clark, arXiv:2110.02965.

[83] J. Kunjummen, M. C. Tran, D. Carney, and J. M. Taylor, arXiv:2110.03629.

[84] T. Ferguson, U-statistics, lecture notes for Statistics 200B, UCLA, 2003 (unpublished).

[85] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevA.106.012441 for a discussion of the shadow postprocessing methods for the adapted protocol

where each circuit is repeated many times, as well as proofs of bounds on the variance of Rényi entropies estimated using shadow tomography, which contains Ref. [86].

[86] P. Busch, Int. J. Theor. Phys. **30**, 1217 (1991).

[87] J. Haah, A. W. Harrow, Z. Ji, X. Wu, and N. Yu, IEEE Trans. Inf. Theory **63**, 5628 (2017).

[88] IBM quantum, https://quantum-computing.ibm.com/ (2021).

[89] Qiskit textbook, Sec. 5.2, https://qiskit.org/textbook/ch-quantum-hardware/measurement-error-mitigation.html.

[90] S. Chen, W. Yu, P. Zeng, and S. T. Flammia, PRX Quantum **2**, 030348 (2021).

[91] Y.-L. Zhang, Y. Huang, and X. Chen, Phys. Rev. B **99**, 014303 (2019).

[92] Z.-H. Sun, J. Cui, and H. Fan, Phys. Rev. A **104**, 022405 (2021).

[93] B. Yoshida and A. Kitaev, arXiv:1710.03363.

[94] C. Kokail, R. van Bijnen, A. Elben, B. Vermersch, and P. Zoller, Nat. Phys. **17**, 936 (2021).

[95] F. Haake, *Quantum Signatures of Chaos*, Springer Series in Synergetics Vol. 54 (Springer, Berlin, 2010).

[96] L. K. Joshi, A. Elben, A. Vikram, B. Vermersch, V. Galitski, and P. Zoller, Phys. Rev. X **12**, 011018 (2022).

[97] Y. Li, X. Chen, and M. P. A. Fisher, Phys. Rev. B **98**, 205136 (2018).

[98] Y. Li, X. Chen, and M. P. A. Fisher, Phys. Rev. B **100**, 134306 (2019).

[99] Y. Bao, S. Choi, and E. Altman, Phys. Rev. B **101**, 104301 (2020).

[100] M. J. Gullans and D. A. Huse, Phys. Rev. X **10**, 041020 (2020).

[101] M. S. Rudner, N. H. Lindner, E. Berg, and M. Levin, Phys. Rev. X **3**, 031005 (2013).

[102] H. C. Po, L. Fidkowski, T. Morimoto, A. C. Potter, and A. Vishwanath, Phys. Rev. X **6**, 041070 (2016).

[103] B. R. Duschatko, P. T. Dumitrescu, and A. C. Potter, Phys. Rev. B **98**, 054309 (2018).

[104] Z. Gong, L. Piroli, and J. I. Cirac, Phys. Rev. Lett. **126**, 160601 (2021).

[105] M. McGinley, S. Leontica, S. J. Garratt, J. Jovanovic, and S. H. Simon (2022). Zenodo, https://doi.org/10.5281/zenodo.5913167.

[106] B. Collins and P. Śniady, Commun. Math. Phys. **264**, 773 (2006).

[107] S. M. Manjegani, Positivity **11**, 239 (2007).

[108] A. W. Marshall, I. Olkin, and B. C. Arnold, *Inequalities: Theory of Majorization and Its Applications*, Springer Series in Statistics Vol. 143 (Springer, New York, 1979).

[109] M. A. Nielsen and J. Kempe, Phys. Rev. Lett. **86**, 5184 (2001).

[110] T. Hiroshima, Phys. Rev. Lett. **91**, 057902 (2003).

[111] M. Horodecki and P. Horodecki, Phys. Rev. A **59**, 4206 (1999).

[112] M. Horodecki, P. Horodecki, and R. Horodecki, Phys. Rev. Lett. **80**, 5239 (1998).

[113] C. H. Bennett, D. P. DiVincenzo, J. A. Smolin, and W. K. Wootters, Phys. Rev. A **54**, 3824 (1996).

[114] H. Barnum, E. Knill, and M. Nielsen, IEEE Trans. Inf. Theory **46**, 1317 (2000).

[115] A. Peres, Phys. Rev. Lett. **77**, 1413 (1996).

[116] W. Dür, J. I. Cirac, M. Lewenstein, and D. Bruß, Phys. Rev. A **61**, 062313 (2000).