Retrieving information from a black hole using quantum machine learning

Lorenzo Leone [1,*] Salvatore F. E. Oliviero [1] Stefano Piemontese [1] Sarah True,[1] and Alioscia Hamma [1,2]

[1]*Physics Department, University of Massachusetts, Boston, Massachusetts 02125, USA*

[2]*Dipartimento di Fisica Ettore Pancini, Università degli Studi di Napoli Federico II, Via Cinthia 80126, Napoli, Italy*

In a seminal paper [P. Hayden and J. Preskill, J. High Energy Phys. 09 (2007) 120], Hayden and Preskill showed that information can be retrieved from a black hole that is sufficiently scrambling, assuming that the retriever has perfect control of the emitted Hawking radiation and perfect knowledge of the internal dynamics of the black hole. In this paper, we show that for t-doped Clifford black holes—that is, black holes modeled by random Clifford circuits doped with an amount t of non-Clifford resources—an information retrieval decoder can be learned with fidelity scaling as $\exp(-\alpha t)$ using quantum machine learning while having access only to out-of-time-order correlation functions. We show that the crossover between learnability and nonlearnability is driven by the amount of nonstabilizerness present in the black hole and sketch a different approach to quantum complexity.

I. INTRODUCTION

The onset of chaos is at the root of the explanation of thermalization in closed quantum systems [1–5]. Although the precise definition of quantum chaos remains elusive, one can usefully refer to it as a bundle of features comprising information scrambling [6–10], a complex pattern of entanglement [11–15], universal behavior of out-of-time-order correlation functions (OTOCs) [16–27], and quantum dynamics that can be efficiently modeled by random unitary operators [18,28–31]. Information scrambling is the quantum analog of the butterfly effect, signaling that local disturbances are spread through operator growth [32–34].

In the context of black-hole physics, one wonders whether the information is destroyed by a black hole or can be recovered from Hawking radiation as the black hole evaporates. It is commonly assumed that a black hole thermalizes information quickly [1,2,35] and that information is also rapidly spread across every part of the system, making the black hole a fast scrambler [36–39]. Moreover, one assumes evolution to be described by a unitary operator U. Under these assumptions, Hayden and Preskill showed that the Hawking radiation releases information very quickly and that one can therefore recover any information initially dumped into the black hole with just a slight overhead of information readout from the subsequent Hawking radiation [40]. This remarkable result hinges on the fact that it is the very scrambling dynamics of the black hole that allows information to be transferred to the Hawking radiation. Yoshida and Kitaev showed how this information can be retrieved by an observer with perfect knowledge of both the initial state of the black hole *and* its unitary dynamics U [41].

Obtaining perfect knowledge of a black hole's internal structure and dynamics sounds like an impossibly daunting task. Could we perhaps learn U by feeding the Hawking radiation into a suitable quantum machine learning (QML) algorithm [42–61]? Such an approach seems promising at first glance, but quickly becomes futile in practice; in order for the recovery algorithm to work, the black hole must be scrambling, but the supposed complexity of scrambling dynamics hinders our ability to learn about its details. Indeed, extensive analysis [62–71] of the barren plateau problem, i.e., system size-exponential vanishing of cost function gradients in variational quantum algorithms, has shown that no QML training protocol can effectively learn U if it is scrambling.

And yet, even so, things are not hopeless. We cannot learn U, but we can train a quantum circuit to decode it by learning a model unitary V that is good enough to unscramble the Hawking radiation, recover the information tossed into the black hole, and perform teleportation. Despite being very different from the original scrambler U, this so-called mocking black hole V is optimized to perform the desired tasks.

In this paper, we show that if a black hole is modeled by a unitary U_t consisting of an element of the Clifford group with a doping t of non-Clifford resources, one can use a cost function directly obtained from the OTOCs to learn a mocking V capable of recovering information from the black hole with fidelity $\mathcal{F}(V) \sim \exp(-\alpha t)$. This is possible because, while the Clifford group is a good scrambler [72–74], it does not produce a complex pattern of entanglement across the subsystems [75,76]. We also show that, with the same technique, one can teleport a state with zero initial knowledge of black-hole dynamics.

II. NOTATION AND TECHNIQUES

Let us introduce some notations and tools that are used throughout the paper. Consider the Hilbert space of n qubits
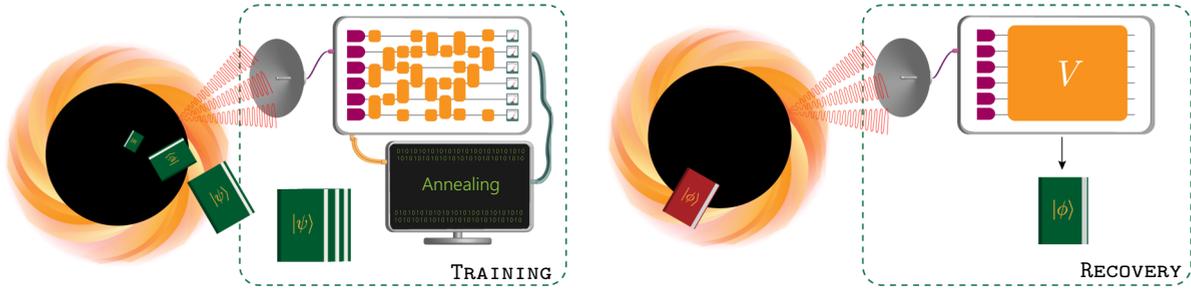
*Lorenzo.Leone001@umb.edu

$\mathcal{H}$, partitioned as $n = n_A + n_B = n_C + n_D$. Let $d_\Lambda \equiv 2^{n_\Lambda}$ with $\Lambda \in \{A, B, C, D\}$. Denote $\mathcal{P}_{n_\Lambda}$ the Pauli group on n_Λ qubits, being $\Lambda \in \{A, B, C, D\}$. Let $P_\Lambda \in \mathcal{P}_\Lambda$ be local Pauli operators acting nonidentically only on Λ, where we omit the identity on the complement of Λ. Let O be an operator on $\mathcal{H}$ and let $P_\Lambda O P_\Lambda$ be the twirling of O with P_Λ; denoting the average over the Pauli group $\mathcal{P}_{n_\Lambda}$ as $\langle P_\Lambda O P_\Lambda \rangle_{\mathcal{P}_{n_\Lambda}} \equiv d_\Lambda^{-2} \sum_{P_\Lambda \in \mathcal{P}_{n_\Lambda}} P_\Lambda O P_\Lambda$, we have

$$\langle P_\Lambda O P_\Lambda \rangle_{P_\Lambda} = \frac{\text{tr}_\Lambda(O)}{d_\Lambda}, \tag{1}$$

where $\text{tr}_\Lambda(\cdot)$ denotes the partial trace on Λ. Consider a Hilbert space $\mathcal{H}_R$ of n_A auxiliary qubits; then the Einstein-Podolsky-Rosen (EPR) pair between A and R is denoted as $|RA\rangle$, living on $\mathcal{H}_R \otimes \mathcal{H}_A$, and is defined as

$$|RA\rangle \equiv \frac{1}{\sqrt{d_A}} \sum_{i=1}^{d_A} |i\rangle_R \otimes |i\rangle_A. \tag{2}$$

Denote $\Pi_{\Lambda\Lambda'} \equiv |\Lambda\Lambda'\rangle \langle \Lambda\Lambda'|$ the projector onto the EPR pair $|\Lambda\Lambda'\rangle$. In what follows, we adopt the following pictorial representation for EPR pairs:

$$|RA\rangle = {}^{R}\!\!\lfloor\!\!\underset{\bullet}{\quad}\!\!\rfloor^{A}, \tag{3}$$

where the horizontal lines represent the EPR pair and the dot captures the normalization $d_A^{-1/2}$. Let O_A be a unitary operator on $\mathcal{H}_A$; then $O_A |RA\rangle = O_R^T |RA\rangle$, where T is the transposition. Diagrammatically

$$\tag{4}$$

Consider two EPR pairs $|RA\rangle |BB'\rangle$, where B' is a n_B qubits auxiliary system, and consider a unitary operator O_{AB} acting on A and B, as $O_{AB} |AR\rangle |BB'\rangle$. The diagrammatic representation corresponding to the projection of $O_{AB} |AR\rangle |BB'\rangle$ onto $\langle RA|$ is

$$\tag{5}$$

Noting that $\langle RA| O_{AB} |RA\rangle |BB'\rangle = d_A^{-1}(\text{tr}_A O_{AB}) |BB'\rangle$, we can use the identity in Eq. (1) to establish the following relation between diagrams:

$$\tag{6}$$

where we considered the diagram corresponding to the projector $d_A^{-2}(\text{tr}_A O_{AB}) |BB'\rangle \langle BB'| (\text{tr}_A O_{AB}^\dagger)$. The above relation will be useful to prove the main equations of the paper.

III. DECODING A BLACK HOLE

Let us start by reviewing the decoding algorithm for information retrieval from a scrambling black hole [40,41]. After a long-time evaporation process, the state of the black hole B is maximally entangled with the already emitted Hawking radiation B', which can be accessed by the observer Bob. This state is described by an EPR pair $|BB'\rangle$. Alice possesses a quantum state A maximally entangled with a reference state R, that is an EPR pair $|RA\rangle$. The maximal entanglement between A and R quantifies the information about A possessed by R. This information is lost when the black hole destroys correlations between A and R. The initial state of the whole system is thus described by $|RA\rangle |BB'\rangle$.

Alice tosses her quantum state—depicted as a journal in Fig. 1—into the black hole, which will soon emit new Hawking radiation D and retain internal degrees of freedom labeled by C. So A, B label the internal degrees of freedom of the black hole before Alice tosses her journal into it and B' labels the Hawking radiation already emitted at that point. After tossing the journal, the internal degrees of freedom of the black hole are described by a state on C and the Hawking radiation is given by $D + B'$.

The black-hole dynamics are modeled by a unitary $U = I_{RB'} \otimes U_{AB}$ which must be random enough to scramble information. After the evolution U, the state of the entire system is given by the pure state $|\Psi\rangle_{RB'CD} = I_{RB'} \otimes U_{AB} |RA\rangle |BB'\rangle$ in $RB'CD$, which diagrammatically is given by

$$\tag{7}$$

Notice that the number n of qubits initially internal to the black hole obeys $n \equiv n_A + n_B = n_C + n_D$, so the Hilbert space of the black-hole interior is shrinking as more Hawking radiation is emitted.

The information possessed by Alice in her journal would be recovered by Bob if he could process the Hawking radiation and end up with a state maximally entangled with another reference state R' in his possession. In the case that Alice's state is a pure state $|\psi_A\rangle$, this would amount to teleporting $|\psi_A\rangle$ from A to R'.

As [40] showed, the existence of such a recovery procedure is contingent on the unitary U being scrambling enough. Because of unitarity, correlations can only be transferred, so if U is scrambling enough to destroy correlations between R and C, these must show up somewhere else and can be decoded by a suitable V^*. The information shared between the qubits of Alice R and the internal degrees of freedom of the black hole C is given by the mutual information $I(R|C) := S(\rho_R) + S(\rho_C) - S(\rho_{RC})$, where $S(\rho_\Lambda) \equiv -\text{tr}\rho_\Lambda \log_2 \rho_\Lambda$ is the von Neumann entropy, and $\rho_\Lambda := \text{tr}_{\bar{\Lambda}} |\Psi\rangle \langle \Psi|$ (with $\bar{\Lambda}$ being the complement of Λ) are reduced density matrices for $\Lambda \in \{R, C, RC\}$. One can easily see [6] that for the state $|\Psi\rangle_{RB'CD}$ one has $I(R|C) = n_A + n_C - S(\rho_{RC})$. Define the OTOC as $\Omega(U) := d^{-1} \langle \text{tr}[P_A P_D(U) P_A P_D(U)] \rangle_{P_A, P_D}$, where $P_D(U) \equiv U^\dagger P_D U$; then, the mutual information can be related to the OTOCs $\Omega(U)$, by $S_2(\rho_{RC}) = -\log_2 \frac{d_A}{d_C} \Omega(U)$

FIG. 1. Quantum machine learning protocol for learning the mocking unitary V capable of performing information recovery from Hawking radiation. Using a cost function given by OTOCs, the model learns to recover a set of known training states (the green journals) that are tossed into the black hole. This model unitary V is then used to decode the Hawking radiation emitted after tossing in a new, unknown journal and retrieve the information contained within it.

proven in [6], as

$$I(R|C) \leqslant -\log_2 d_A^2 \Omega(U). \tag{8}$$

A unitary U is said to be *scrambling* if and only if $\Omega(U) \simeq d_A^{-2} + d_D^{-2} - (d_A d_D)^{-2}$, where $\simeq$ means *up to an order d^{-2}*. In the limit of $d_A \ll d_D$, the scrambling dynamics will therefore imply that $I(R|C) \simeq 0$, thus resulting in the destruction of correlations between R and C. Thanks to the unitarity of U, all the information has been transferred to Bob; the mutual information $I(R|DB')$ between R and the qubit in Bob's possession B' together with the new emitted Hawking radiation D is maximal, i.e.,

$$I(R|DB') \simeq n_A. \tag{9}$$

Let us see how Bob is able to recover the information initially possessed by Alice. Bob possesses an EPR pair $A'R'$ and applies a unitary V^* to the old Hawking radiation B' and one-half of his pair A'. Then, by reading the additional Hawking radiation D, effectively entangling it in another EPR pair DD', Bob projects by $\Pi_{DD'}$ onto DD' to obtain a final state $|\Psi_{\text{out}}\rangle = \frac{1}{\sqrt{P_{\text{out}}}}\Pi_{DD'}V_{B'A'}^*U_{AB}|RA\rangle|BB'\rangle|A'R'\rangle$, where P_{out} is a normalization. Diagrammatically

$$|\Psi_{out}\rangle = \frac{1}{\sqrt{P_{out}}} \quad \raisebox{-1em}{\text{[diagram]}} \quad . \tag{10}$$

Following [41,77,78], the normalization P_{out} can be computed with the help of the diagrammatic formalism introduced in Sec. II as

$$P_{out} \equiv \quad \raisebox{-1em}{\text{[diagram]}} \quad . \tag{11}$$

The above diagram is obtained from contracting the diagram corresponding to $\Pi_{DD'}V_{B'A'}^*U_{AB}|RA\rangle|BB'\rangle|A'R'\rangle$ with itself. By using Eq. (1) and its diagrammatical representation in Eq. (6),

one arrives to

$$P_{\text{out}} = \frac{1}{d}\langle \text{tr}[P_D(U)P_A P_D(V)P_A]\rangle_{P_A,P_D}. \tag{12}$$

The recovery algorithm is successful if one obtains a maximally entangled pair between R and R' that is factorized from the rest. In this way, the information about A originally in the hands of Alice has been successfully transferred to Bob [40]. In other words, the final state must read $|\Psi_{\text{out}}\rangle \simeq |RR'\rangle \otimes |\text{rest}\rangle_{CC'} \otimes |DD'\rangle$. Success of the algorithm can be established by computing the fidelity $\mathcal{F}(V) := \text{tr}(\Psi_{\text{out}}\Pi_{RR'})$, where $\Psi_{\text{out}} \equiv |\Psi_{\text{out}}\rangle\langle\Psi_{\text{out}}|$; the product $\mathcal{F}(V)P_{\text{out}}$ can be easily computed using diagrammatical techniques as

$$\mathcal{F}(V)P_{out} = \quad \raisebox{-1em}{\text{[diagram]}} \quad , \tag{13}$$

which is obtained by crunching the diagram in Eq. (10) with the one corresponding to $|RR'\rangle$ [see Eq. (3)]. Using Eq. (6) one arrives to

$$\mathcal{F}(V)P_{\text{out}} = \frac{1}{d_A^2 d}\langle \text{tr}P_D(U)P_D(V)\rangle_{P_D}. \tag{14}$$

Putting together Eqs. (12) and (14), one derives the main equation of the paper, i.e., the recovery fidelity $\mathcal{F}(V)$ as a function of the scrambler U and the decoder V:

$$\mathcal{F}(V) = \frac{1}{d_A^2}\frac{\langle \text{tr}[P_D(U)P_D(V)]\rangle_{P_D}}{\langle \text{tr}[P_D(U)P_A P_D(V)P_A]\rangle_{P_A,P_D}}. \tag{15}$$

In the above expression, the role of V is to *mock* the behavior of the black hole modeled by U. If $V = U$, the behavior of U is obviously replicated perfectly. Indeed, the result in [41] shows that in the ideal case $V = U$, the fidelity $\mathcal{F}(U)$ reads $\mathcal{F}(U) = \frac{1}{d_A^2\Omega(U)}$, and one can see that if the black hole is indeed scrambling, one obtains a fidelity $\mathcal{F}(U) = 1 - O(d_A^2/d_D^2)$. One can easily see [77] that $\text{tr}[P_D(U)P_A P_D(V)P_A] = \langle\Psi_{\text{out}}^{P_A}|P_D \otimes P_{D'}^*|\Psi_{\text{out}}^{P_A}\rangle$ where $|\Psi_{\text{out}}^{P_A}(U,V)\rangle := (U_{AB} \otimes V_{B'A'}^*)(P_A \otimes$

$1_{A'RR'BB'}$) $|RA\rangle$ $|BB'\rangle$ $|A'R'\rangle$. Thus, $\mathcal{F}(V)$ is a quantity that can be recast as a ratio of expectation values of local observables (on D) with supports on accessible parts of the system, D being the emitted Hawking radiation. Note that, if the qubits in A are $n_A = O(1)$, the new emitted Hawking radiation D that is collected by Bob must obey $n_D = n_A + \log_2 \epsilon^{-1/2}$ to feature a ϵ-maximal mutual information $I(R|DB') = n_A - \epsilon$ [see Eq. (8)]. Thus, collecting slightly more qubits than the ones contained in A, Bob decodes the scrambled information by local $O(1)$ measurements.

A. Learning the black-hole decoder

Can we learn the black hole by observing its Hawking radiation by some quantum machine learning algorithm? The presence of barren plateaus [62–69] seems to forbid any kind of learning if the black hole is scrambling enough, which is also the condition that allows information retrieval. The main result of this paper is that—under suitable conditions- - one can nevertheless learn a decoder V that mocks the behavior of the black hole sufficiently enough to recover information.

Consider the case that the black hole is modeled by a t-doped random Clifford circuit $U \in \mathcal{C}_t :=$ $\{C_{t-1}T_t C_{t-2} \cdots T_2 C_1 T_1 C_0 \,|\, C_k \in \mathcal{C}(2^n)\}$ with T_k being single qubit T gates applied on random qubits [75]. Such operators are scrambling enough [75,79] for every t. However, for $t = 0$, i.e., random Clifford operators, a learning protocol based on a METROPOLIS algorithm is possible [11,80,81]. Moreover, it has been recently shown that low t-doping random Clifford circuits can be efficiently disentangled by annealing through a suitable METROPOLIS algorithm [15]. In the following, we show a similar METROPOLIS algorithm based on the cost function $\mathcal{F}(V) \equiv [1 + c(U, V)]^{-1}$ to learn the mocking operator V.

From the quantum machine learning point of view, we describe the procedure as follows. To learn V, we first train our estimator $c(U, V)$ by preparing a number $O(n^2)$ of copies of a known state. This state is used as the input of a black hole. Starting with the identity $V_0 = I$, one tries the cost function $c(U, V)$ and optimizes it by local modifications using a Clifford gate g, so that $V_0 \mapsto gV_0$. Acceptance of the gate depends on improving the cost function $c(U, V)$ or lowering it with Boltzmann probability parametrized by β (see Appendix B). The many copies of the training journal tossed in the black hole (Fig. 1) correspond to the steps of the METROPOLIS algorithm needed. After the training, one has settled on a V that minimizes the cost function $c(U, V)$ with the number of steps allowed. One can then use V in the recovery algorithm, and recover information about a new journal tossed into U with fidelity $\mathcal{F}(V)$.

The numerical simulation of the whole algorithm is computationally very expensive. In order to perform the numerical simulations on a smaller space, we can exploit the fact that we are working in the scenario $d_C \ll d_D$. By averaging over the Pauli group on C instead of D, the cost function $c(U, V)$ can be computed as

$$c(U, V) = \frac{\sum_{P_C, P_A \neq 1_A} |\text{tr}[(U^\dagger P_C V P_A)]|^2}{\sum_{P_C} |\text{tr}[(U^\dagger P_C V)]|^2}, \qquad (16)$$
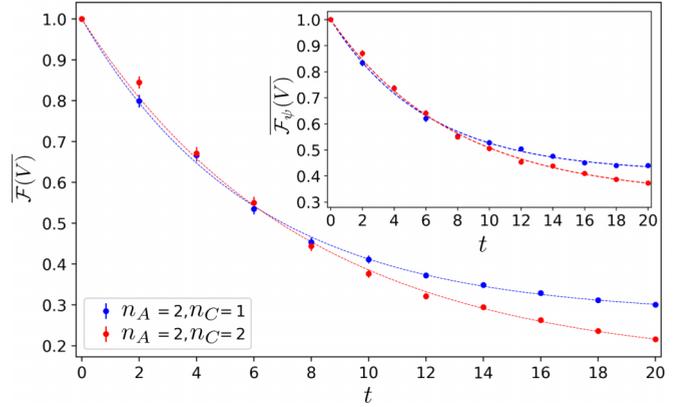

FIG. 2. Average over 200 realizations of the fidelity $\mathcal{F}(V)$, denoted as $\overline{\mathcal{F}(V)}$, at the end of the METROPOLIS algorithm as a function of the doping t, fit to $a \exp(-\alpha t) + b$ where $\alpha = 0.167$, $a = 0.72$, $b = 0.28$ for $n_C = 1$, while $\alpha = 0.129$, $a = 0.85$, $b = 0.15$ for $n_C = 2$. Each realization is performed for a newly generated black-hole unitary U. Inset: Same for the teleportation fidelity $\mathcal{F}_\psi(V)$, with the same values for α (see Appendix B).

which can be proven using Eq. (1) (see Appendix B). The simulations are conducted for a system of $n = 10$ qubits initialized in the state $|0\rangle^{\otimes n}$, with $n_A = 2$ and $n_C = 1, 2$. U is a random Clifford circuit consisting of t layers each with $O(n^2)$ local gates controlled-NOT (CNOT), H, and P interspersed by a single T gate per layer [15,75]. We indeed stress that $O(n^2)$ gates from the local gate set $\{\text{CNOT}, H, P\}$ are sufficient to distill any Clifford operator [82]. This fact lies behind the reason why, during the training process, we make use only of $O(n^2)$ state preparations that result in $O(n^2)$ many updates $V \mapsto gV$.

The results are shown in Fig. 2. As we can see, for $t = 0$ a mocking operator V for a black hole modeled by a pure random Clifford circuit can be learned with perfect fidelity. As the number t of non-Clifford resources increases, the fidelity for the mocking operator V decreases exponentially in t. Notice that the fidelity also depends on the size n_C of the interior of the black hole. The fact that we are not learning the operator U can be checked by comparing V with U. Indeed, this would be forbidden by the barren plateau result found in [62]. We find that $d^{-1}|\text{tr}U^\dagger V|^2 < 0.004$ for every value of t, including $t = 0$. It is important to remark that we only try to reconstruct the mocking V using merely Clifford resources, in spite of the fact that the original U also contains non-Clifford resources. Although searching for V within the Clifford group means we cannot reconstruct perfectly U; the learning algorithm fails with the addition of non-Clifford resources. For instance, for $t = 6$, $n_A = 2$, $n_C = 1$, the fidelity obtained using only Clifford gates attains a value of 0.53 versus 0.4 when allowing the inclusion of T gates. Universal resources are powerful, but pollute the algorithm [15].

The same recovery algorithm of [77] can be used to employ the recovery unitary V to perform quantum teleportation between Alice and Bob. We now show that the previous learning of the mocking V can be efficiently used to perform teleportation without having any previous knowledge of U. Let $|\psi_A\rangle$ be the pure state of Alice to be teleported, and let $|\Psi_{\text{out}}^{\psi_A}\rangle$ be

the output state of the black hole manipulated by the decoding protocol employed by Bob, i.e.,

$$\left|\Psi_{\text{out}}^{\psi_A}\right\rangle := \frac{1}{\sqrt{P_\psi}} \Pi_{DD'} V_{B'A'}^* U_{AB} \left|RA\right\rangle \left|BB'\right\rangle \left|\psi_A\right\rangle, \qquad (17)$$

where P_ψ is a normalization. Similarly to Eq. (15), the fidelity $\mathcal{F}_\psi(V)$ to teleport a state $|\psi_A\rangle$, i.e., $\mathcal{F}_\psi(V) \equiv \text{tr}(\psi_A |\psi_{\text{out}}^{\psi_A}\rangle \langle \psi_{\text{out}}^{\psi_A}|)$, can be computed as (see Appendix A)

$$\mathcal{F}_\psi(V) = \frac{\langle \text{tr}[\psi_A \text{tr}_B(V^\dagger P_C U)] \text{tr}[\psi_A \text{tr}_B(U^\dagger P_C V)] \rangle_{P_C}}{\langle \text{tr}[\psi_A \text{tr}_B(V^\dagger P_C U) \text{tr}_B(U^\dagger P_C V)] \rangle_{P_C}}. \quad (18)$$

In the above formulas we denoted $\psi_A \equiv |\psi_A\rangle \langle \psi_A|$. In the inset of Fig. 2 we plot the value $\mathcal{F}_\psi$ as a function of the doping t. Again, perfect teleportation is achieved for a Clifford black hole ($t = 0$) while the fidelity $\mathcal{F}_\psi(V)$ decreases as non-Clifford resources are employed.

B. Quantum complexity

We have seen that the Clifford black hole can thus be perfectly learned—again, in the sense that one can learn the mocking operator V—while this learning becomes less and less reliable with the injection of non-Clifford resources. Why is that? This is another instance of the fact [75,83] that quantum complexity arises when scrambling (that is, efficient entanglement) is combined with nonstabilizerness, or *magic* [83–88]: the resource that is at the root of quantum advantage for quantum computers and the nonsimulability of generic quantum systems by classical computers [82,89].

While Clifford circuits are just as efficient in scrambling as a Haar-random unitary, they do not create a complex pattern of entanglement, and the fluctuations of entanglement are very different [75]. Quantum complexity is driven by the conspiracy of entanglement and nonstabilizerness (magic), or, in other words, by the complexity of entanglement [11,12,90]. In particular, the ensemble fluctuations of the OTOCs defined as $\Delta\Omega_t := \langle \Omega(U)^2 \rangle_{U \in \mathcal{C}_t} - \langle \Omega(U) \rangle_{U \in \mathcal{C}_t}^2$ behave very differently, and it is not surprising that they are governed by the eight-point OTOC, which probes more fine-grained properties of scrambling [75] (see Appendix C):

$$\Delta\Omega_t \simeq \frac{1}{d_A^2 d_D^2} [\langle \text{OTOC}_8(U) \rangle_{U \in \mathcal{C}_t} + O(d^{-2})]. \qquad (19)$$

Using the techniques introduced in [75,76], one can compute Eq. (19) and find $\Delta\Omega_t \simeq d_A^{-2} d_D^{-2} (\frac{3}{4})^t$, which interpolates between $\Delta\Omega_t = O(d_D^{-2})$ and $O(d_D^{-2} d^{-2})$ for $t = O(1), O(n)$,

respectively. The relative fluctuations for the cost function $\mathcal{F}^{-1}$ for $U = V$ are therefore very small, revealing a barren plateau. However, in searching for the mocking black hole V we minimize the cost function for a V that is far from the ideal U. The situation here is similar to the effectiveness of the success of the disentangling algorithm by the METROPOLIS algorithm [11,15] in random circuits that are at 2-designs. A formal proof of why this kind of algorithms does converge in polynomial time is beyond the scope of this paper.

IV. CONCLUSIONS

A black hole is a very mysterious object that crunches and scrambles information. The internal dynamics of the black hole cannot be resolved even by a quantum machine learning algorithm, so information retrieval from a black hole through Hawking radiation seems hopeless. However, we have shown that one can learn a mocking unitary that is capable of unscrambling and decoding the Hawking radiation by a METROPOLIS algorithm based on the observation of out-of-time-order correlation functions. The learning is possible for black holes that can be modeled by slightly polluted Clifford circuits. Highly polluted black holes cannot be learned. This result illustrates the crossover from simpler to complex quantum behavior and demonstrates how the onset of quantum chaos is driven by the conjunction of entanglement and nonstabilizerness. In view of the results on quantum certification in [91], it would be interesting to show that the same intractability also shows up in the number of measurements needed to reliably measure the OTOCs in order to effectively perform information retrieval.

APPENDIX A: PROOF OF EQ. (18)

To prove Eq. (18) we can proceed using the techniques introduced in the main text, although the setup is slightly different. First, consider the initial state (output of the black-hole evolution) to be $|\Phi\rangle_{B'CD} := U_{AB} |\psi\rangle_A |BB'\rangle$. Bob first applies the decoder unitary $V_{B'A'}^*$ on the state $|\Phi\rangle_{B'CD} |A'R'\rangle$ and then projects it on $\Pi_{DD'}$. The output state reads

$$|\Phi_{out}\rangle = \frac{1}{\sqrt{P_\psi}} \Pi_{DD'} V_{B'A'}^* |\Phi\rangle_{B'CD} |A'R'\rangle = \frac{1}{\sqrt{P_\psi}}$$

$$, \qquad (A1)$$

where P_ψ is a normalization factor. Using the diagrammatical representation of $|\Phi_{\text{out}}\rangle$, one computes

$$\mathcal{F}_\psi(V)P_\psi = \qquad = \frac{1}{d_B} \left\langle \text{tr}[\psi_A \, \text{tr}_B(V^\dagger P_C U)] \, \text{tr}[\psi_A \, \text{tr}_B(U^\dagger P_C V)] \right\rangle_{P_C}, \qquad (A2)$$

and

$$P_\psi \equiv \qquad = \frac{1}{d_B} \left\langle \text{tr}[\psi_A \, \text{tr}_B(V^\dagger P_C U) \, \text{tr}_B(U^\dagger P_C V)] \right\rangle_{P_C}, \qquad (A3)$$

and taking the ratio between Eqs. (A2) and (A3) one obtains Eq. (18).

APPENDIX B: METROPOLIS ALGORITHM

First let us write Eq. (2) in a smaller space, proving Eq. (3). Consider the denominator of Eq. (2); using $\text{tr}(\mathcal{O}^2) \equiv \text{tr}(T\mathcal{O}^{\otimes 2})$ where T is the swap operator defined on $\mathcal{H}^{\otimes 2}$, we can rewrite this term as

$$\langle \text{tr}[P_D(U)P_A P_D(V)P_A] \rangle_{P_A, P_D}$$
$$= \left\langle \text{tr}(U^\dagger \otimes V^\dagger P_D^{\otimes 2} U \otimes V P_A^{\otimes 2} T) \right\rangle_{P_D, P_A}$$
$$= \frac{1}{d_A d_D} \text{tr}(U^\dagger \otimes V^\dagger T_D U \otimes V T_A T)$$
$$= \frac{1}{d_D d_A} \text{tr}(V^\dagger \otimes U^\dagger T_C U \otimes V T_A)$$
$$= \frac{1}{d d_A^2} \sum_{P_C, P_A} |\text{tr}(U^\dagger P_C V P_A)|^2, \qquad (B1)$$

where we used the following facts: $\sum_{P_\Lambda} P_\Lambda^{\otimes 2} = d_\Lambda T_\Lambda$ where T_Λ is the swap operator having support on Λ and $T T_\Lambda = T_{\bar\Lambda}$ where $\bar\Lambda$ is the complement of Λ. Similarly for the numerator of Eq. (2)

$$\langle \text{tr}[P_D(U)P_D(V)] \rangle_{P_D} = \frac{1}{d} \sum_{P_C} |\text{tr}(U^\dagger P_C V)|^2. \qquad (B2)$$

Putting it all together, we find

$$\mathcal{F}(V) = [1 + c(U, V)]^{-1}, \qquad (B3)$$

where $c(U, V)$ is defined in Eq. (3). Note that $c(U, V) \geqslant 0$ with $c(U, V) = 0$ if and only if $\mathcal{F}(V) = 1$, making $c(U, V)$ the best candidate for a cost function. Let us describe the

algorithm to numerically find an optimal recovery mocking unitary V.

Algorithm 1 Training Algorithm.

1:   $U \leftarrow$ random doped Clifford circuit $\mathcal{C}_t$
2:   $\mathcal{F}_{\text{max}} \leftarrow \mathcal{F}(U)$
3:   $V \leftarrow \mathbb{1}$
4:   $\mathcal{F} \leftarrow \mathcal{F}(V)$
5:   **while** $\mathcal{F} < \mathcal{F}_{\text{max}}$ **do**
6:      $V_{old} \leftarrow V$
7:      $\mathcal{F}_{old} \leftarrow \mathcal{F}$
8:      $g \leftarrow$ random Clifford gate
9:      $V \leftarrow gV$
10:     $\mathcal{F} \leftarrow \mathcal{F}(V)$
11:     **if** $\mathcal{F} < \mathcal{F}_{old}$ **then**
12:        $r \in [0, 1]$
13:        **if** $r < \exp[-\beta(\mathcal{F}^{-1} - \mathcal{F}_{old}^{-1})]$ **then**
14:          Undo change: $V \leftarrow V_{old}$
15:          $\mathcal{F} \leftarrow \mathcal{F}_{old}$
16:        **end if**
17:     **end if**
18: **end while**

A constant cooling schedule is employed with $\beta = 250$. In each iteration of the *while* loop, several $2^n \times 2^n$ matrix multiplications must be performed to obtain the operator $U P_C V^\dagger P_A$. These operations quickly become computationally expensive as we increase the system size; even the fastest algorithms for this task [92,93] would require a computation time that

TABLE I. The table shows the fit parameters for the exponential model applied to the data. The model equation is $a \exp(-\alpha t) + b$, where α, a and b are the fit parameters, while t is the number of t-gates polluting the Clifford circuit. The parameters are computed for both $\overline{\mathcal{F}}_f^{(U,V)}$ and $\overline{\mathcal{F}}_\psi^{(U,V)}$ for $n_C = 1, 2$.

		$n_C = 1$	$n_C = 2$
	α	0.167	0.129
$\overline{\mathcal{F}}_f^{(U,V)}$	a	0.7243	0.8483
	b	0.2757	0.1517
$\overline{\mathcal{F}}_\psi^{(U,V)}$	a	0.5860	0.6794
	b	0.4140	0.3206

grows exponentially in n. Furthermore, to avoid indefinite runtimes when the gradient of $c(U, V)$ becomes small, we halt the algorithm after $T_{\max}$ steps. The value of $T_{\max}$ used is $100n^2$ when $n_C = 1$ and $300n^2$ when $n_C = 2$, chosen empirically by observing the algorithm's runtime when $t = 0$. Note that the time requirement grows with the number of qubits as $T_{\max} \propto n^2$, as seen also in [11,15,80]. Therefore, the numerical analysis in this paper is performed for systems of ten qubits, as larger system sizes would require exponentially larger computational times.

The parameters for the results fit to $a \exp(-\alpha t) + b$ are shown in Table I.

APPENDIX C: THE COMPLEXITY OF SCRAMBLING: THE EIGHT-POINT OTOC

In this section, we present a single correlation function that tells us whether the algorithm is going to fail or not, i.e., the eight-point out-of-time-order correlator, proving Eq. (5). First, let us write the OTOC $\Omega(U) = \langle \text{tr}[P_A P_D(U) P_A P_D(U)] \rangle_{P_A, P_D}$ as

$$\Omega(U) = \frac{1}{d_A^2} + \frac{1}{d_D^2} - \frac{1}{d_A^2 d_D^2} + \frac{1}{d_A^2 d_D^2}$$
$$\times \sum_{P_A, P_D \neq 1} \frac{1}{d} \text{tr}[P_A P_D(U) P_A P_D(U)] \quad \text{(C1)}$$

and we thus can alternately define a scrambling unitary U as one such that

$$f(U) \equiv \frac{1}{d_A^2 d_D^2} \sum_{P_A, P_D \neq 1} \text{otoc}_4(P_A, P_D(U)) = O(d^{-1}), \quad \text{(C2)}$$

where we defined $\text{OTOC}_4(P_A, P_D(U)) := \frac{1}{d} \text{tr}[P_A P_D(U) P_A P_D(U)]$. Note that for $U \in \mathcal{C}(2^n)$ being a Clifford operator $\text{OTOC}_4(P_A, P_D(U)) = \pm 1$ for any choice of

P_A, P_D, but $f(U) = O(d^{-1})$ still holds for Clifford unitaries, i.e., Clifford unitaries are scramblers. Note also that

$$\langle f(U) \rangle_U = \frac{(d_A^2 - 1)(d_D^2 - 1)}{d_A^2 d_D^2} \langle \text{OTOC}_4(P_1, P_2(U)) \rangle_U, \quad \text{(C3)}$$

for any P_1, P_2 chosen to be two nonidentity Pauli operators and U being a Clifford or t-doped Clifford operator. Now let us look at the fluctuations in the ensemble of t-doped Clifford operators $U \in \mathcal{C}_t$. Direct calculation leads to

$$\Delta \Omega_t = \langle f(U)^2 \rangle_U - \langle f(U) \rangle_U^2 = \langle f(U)^2 \rangle_U + O(d^{-2}), \quad \text{(C4)}$$

and the second equality follows from the definition of scrambling unitaries, i.e., $\langle f(U) \rangle_U = O(d^{-1})$ [see Eq. (C2)], and it holds for any $t \in \mathbb{N}$. Thus, turn to analyze $\langle f(U)^2 \rangle_U$:

$$\langle f(U)^2 \rangle_U = \frac{(d_A^2 - 1)(d_D^2 - 1)}{d_A^4 d_D^4} \langle \text{OTOC}_4(P_1, P_2(U))^2 \rangle_U$$
$$+ \langle R(U) \rangle_U, \quad \text{(C5)}$$

$$R(U) := \frac{1}{d_A^4 d_D^4} \sum_{P_A, P_D, P_A', P_D' \neq 1} [1 - \delta(P_A = P_A')\delta(P_D = P_D')]$$
$$\times \text{OTOC}_4(P_A, P_D(U)) \text{OTOC}_4(P_A', P_D'(U)). \quad \text{(C6)}$$

Using standard techniques of the Haar measure over groups [75,94–96] one proves that for any $t \in \mathbb{N}$ we have $\langle R(U) \rangle = O(d^{-2})$. Lastly, using the fact that the Pauli group forms a 1-design [18], it is easy to prove [18] the following:

$$\text{otoc}_4(P_1, P_2(U))^2 = \text{OTOC}_8(U), \quad \text{(C7)}$$

where we defined the following eight-point out-of-time-order correlation function:

$$\text{OTOC}_8(U)$$
$$:= \frac{1}{d} \langle \text{tr}[P_1 P_2(U) P_1 P_2(U) P P_1 P_2(U) P_1 P_2(U) P] \rangle_{P \in \mathcal{P}_n}, \quad \text{(C8)}$$

where P_1 and P_2 are any nonidentity Pauli operators. We can finally write the following formula:

$$\Delta \Omega_t = \frac{(d_A^2 - 1)(d_D^2 - 1)}{d_A^4 d_D^4} \langle \text{OTOC}_8(U) \rangle_U + O(d^{-2}), \quad \text{(C9)}$$

and in the limit of $d_D \gg d_A \gg 1$ one arrives to Eq. (5). Thus the ensemble fluctuations of the four-point OTOC are proportional to the ensemble average of an eight-point correlation function, which probes more fine-grained properties of scrambling, i.e., the complexity of scrambling. We thus define the complexity of scrambling produced by some unitary operator U as the behavior of the eight-point correlation function defined in Eq. (C8).

[1] S. Lloyd, Black Holes, Demons and the loss of coherence: How complex systems get information, and what they do with it, Ph.D. thesis, Rockefeller University, 1988.

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

[3] N. Linden, S. Popescu, A. J. Short, and A. Winter, Phys. Rev. E **79**, 061103 (2009).

[4] S. Popescu, A. J. Short, and A. Winter, Nat. Phys. **2**, 754 (2006).

[5] M. Rigol and L. F. Santos, Phys. Rev. A **82**, 011604(R) (2010).

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

[7] D. Ding, P. Hayden, and M. Walter, J. High Energy Phys. 12 (2016) 145.

[8] W. Brown and O. Fawzi, Scrambling speed of random quantum circuits, arXiv:1210.6644.

[9] B. Yan, L. Cincio, and W. H. Zurek, Phys. Rev. Lett. **124**, 160603 (2020).

[10] A. Touil and S. Deffner, Quantum Sci. Technol. **5**, 035005 (2020).

[11] C. Chamon, A. Hamma, and E. R. Mucciolo, Phys. Rev. Lett. **112**, 240501 (2014).

[12] Z.-C. Yang, A. Hamma, S. M. Giampaolo, E. R. Mucciolo, and C. Chamon, Phys. Rev. B **96**, 020408(R) (2017).

[13] Z.-W. Liu, S. Lloyd, E. Y. Zhu, and H. Zhu, Phys. Rev. Lett. **120**, 130502 (2018).

[14] Z.-W. Liu, S. Lloyd, E. Zhu, and H. Zhu, J. High Energy Phys. 07 (2018) 041.

[15] S. True and A. Hamma, Quantum **6**, 818 (2022).

[16] A. I. Larkin and Y. N. Ovchinnikov, J. Exp. Theor. Phys. **28**, 1200 (1969) [Sov. Phys.-JETP **20**, 762 (1965)].

[17] A. Kitaev, Hidden correlations in the Hawking radiation and thermal noise, Talk Given at the Fundamental Physics Prize Symposium (2014), Vol. 10.

[18] D. A. Roberts and B. Yoshida, J. High Energy Phys. 04 (2017) 121.

[19] D. J. Luitz and Y. Bar Lev, Phys. Rev. B **96**, 020406(R) (2017).

[20] K. Hashimoto, K. Murata, and R. Yoshii, J. High Energy Phys. 10 (2017) 138.

[21] N. Yunger Halpern, Phys. Rev. A **95**, 012120 (2017).

[22] I. García-Mata, M. Saraceno, R. A. Jalabert, A. J. Roncaglia, and D. A. Wisniacki, Phys. Rev. Lett. **121**, 210601 (2018).

[23] T. Rakovszky, F. Pollmann, and C. W. von Keyserlingk, Phys. Rev. X **8**, 031058 (2018).

[24] C.-J. Lin and O. I. Motrunich, Phys. Rev. B **97**, 144304 (2018).

[25] F. Borgonovi, F. M. Izrailev, and L. F. Santos, Phys. Rev. E **99**, 052143 (2019).

[26] E. M. Fortes, I. García-Mata, R. A. Jalabert, and D. A. Wisniacki, Phys. Rev. E **100**, 042201 (2019).

[27] L. Leone, S. F. E. Oliviero, and A. Hamma, Entropy **23**, 1073 (2021).

[28] W. G. Brown, Random quantum dynamics: From random quantum circuits to quantum chaos, Ph.D. thesis, Dartmouth College, 2010.

[29] W. G. Brown and L. Viola, Phys. Rev. Lett. **104**, 250501 (2010).

[30] A. Harrow and S. Mehraban, arXiv:1809.06957.

[31] H. Gharibyan, M. Hanada, S. H. Shenker, and M. Tezuka, J. High Energy Phys. 07 (2018) 124.

[32] D. A. Roberts and D. Stanford, Phys. Rev. Lett. **115**, 131603 (2015).

[33] J. S. Cotler, D. Ding, and G. R. Penington, Ann. Phys. (NY) **396**, 318 (2018).

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

[35] L. D'Alessio, Y. Kafri, A. Polkovnikov, and M. Rigol, Adv. Phys. **65**, 239 (2016).

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

[37] N. Lashkari, D. Stanford, M. Hastings, T. Osborne, and P. Hayden, J. High Energy Phys. 04 (2013) 022.

[38] S. H. Shenker and D. Stanford, J. High Energy Phys. 12 (2014) 046.

[39] J. Maldacena, S. H. Shenker, and D. Stanford, J. High Energy Phys. 08 (2016) 106.

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

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

[42] J. Biamonte, P. Wittek, N. Pancotti, P. Rebentrost, N. Wiebe, and S. Lloyd, Nature (London) **549**, 195 (2017).

[43] C. Ciliberto, M. Herbster, A. D. Ialongo, M. Pontil, A. Rocchetto, S. Severini, and L. Wossnig, Proc. R. Soc. A **474**, 20170551 (2018).

[44] V. Dunjko and H. J. Briegel, Rep. Prog. Phys. **81**, 074001 (2018).

[45] S. Lloyd and C. Weedbrook, Phys. Rev. Lett. **121**, 040502 (2018).

[46] K. Poland, K. Beer, and T. J. Osborne, arXiv:2003.14103.

[47] S. Lloyd, M. Schuld, A. Ijaz, J. Izaac, and N. Killoran, arXiv:2001.03622.

[48] J. R. McClean, J. Romero, R. Babbush, and A. Aspuru-Guzik, New J. Phys. **18**, 023023 (2016).

[49] J. Romero, J. P. Olson, and A. Aspuru-Guzik, Quantum Sci. Technol. **2**, 045001 (2017).

[50] R. LaRose, A. Tikku, É. O'Neel-Judy, L. Cincio, and P. J. Coles, npj Quantum Inf. **5**, 57 (2019).

[51] A. Arrasmith, L. Cincio, A. T. Sornborger, W. H. Zurek, and P. J. Coles, Nat. Commun. **10**, 3438 (2019).

[52] M. Cerezo, A. Poremba, L. Cincio, and P. J. Coles, Quantum **4**, 248 (2020).

[53] K. Sharma, S. Khatri, M. Cerezo, and P. J. Coles, New J. Phys. **22**, 043006 (2020).

[54] C. Bravo-Prieto, R. LaRose, M. Cerezo, Y. Subasi, L. Cincio, and P. J. Coles, arXiv:1909.05820.

[55] M. Cerezo, K. Sharma, A. Arrasmith, and P. J. Coles, npj Quantum Information **8**, 113 (2022).

[56] B. Commeau, M. Cerezo, Z. Holmes, L. Cincio, P. J. Coles, and A. Sornborger, arXiv:2009.02559.

[57] Y. Li and S. C. Benjamin, Phys. Rev. X **7**, 021050 (2017).

[58] X. Yuan, S. Endo, Q. Zhao, Y. Li, and S. C. Benjamin, Quantum **3**, 191 (2019).

[59] K. Sharma, M. Cerezo, L. Cincio, and P. J. Coles, Phys. Rev. Lett. **128**, 180505 (2022).

[60] T. Volkoff and P. J. Coles, Quantum Sci. Technol. **6**, 025008 (2021).

[61] K. Endo, T. Nakamura, K. Fujii, and N. Yamamoto, Phys. Rev. Res. **2**, 043442 (2020).

[62] Z. Holmes, A. Arrasmith, B. Yan, P. J. Coles, A. Albrecht, and A. T. Sornborger, Phys. Rev. Lett. **126**, 190501 (2021).

[63] J. R. McClean, S. Boixo, V. N. Smelyanskiy, R. Babbush, and H. Neven, Nat. Commun. **9**, 4812 (2018).

[64] M. Cerezo, A. Sone, T. Volkoff, L. Cincio, and P. J. Coles, Nat. Commun. **12**, 1791 (2021).

[65] S. Wang, E. Fontana, M. Cerezo, K. Sharma, A. Sone, L. Cincio, and P. J. Coles, Nat. Commun. **12**, 6961 (2021).

[66] M. Cerezo and P. J. Coles, Quantum Sci. Technol. **6**, 035006 (2021).

[67] A. Arrasmith, M. Cerezo, P. Czarnik, L. Cincio, and P. J. Coles, Quantum **5**, 558 (2021).

[68] E. Grant, L. Wossnig, M. Ostaszewski, and M. Benedetti, Quantum **3**, 214 (2019).

[69] E. R. Anschuetz and B. T. Kiani, Nat. Commun. **13**, 7760 (2022).

[70] R. J. Garcia, C. Zhao, K. Bu, and A. Jaffe, arXiv:2205.06679.

[71] L. Leone, S. F. E. Oliviero, L. Cincio, and M. Cerezo, arXiv:2211.01477.

[72] A. J. Scott, J. Phys. A: Math. Theor. **41**, 055308 (2008).

[73] H. Zhu, Phys. Rev. A **96**, 062336 (2017).

[74] H. Zhu, R. Kueng, M. Grassl, and D. Gross, arXiv:1609.08172.

[75] L. Leone, S. F. E. Oliviero, Y. Zhou, and A. Hamma, Quantum **5**, 453 (2021).

[76] S. F. Oliviero, L. Leone, and A. Hamma, Phys. Lett. A **418**, 127721 (2021).

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

[78] R. Li and J. Wang, Phys. Rev. D **106**, 046011 (2022).

[79] B. Yoshida, arXiv:2106.15628.

[80] D. Shaffer, C. Chamon, A. Hamma, and E. R. Mucciolo, J. Stat. Mech.: Theory Exp. (2014) P12007.

[81] S. Zhou, Z.-C. Yang, A. Hamma, and C. Chamon, SciPost Phys. **9**, 087 (2020).

[82] S. Aaronson and D. Gottesman, Phys. Rev. A **70**, 052328 (2004).

[83] L. Leone, S. F. E. Oliviero, and A. Hamma, Phys. Rev. Lett. **128**, 050402 (2022).

[84] E. T. Campbell and D. E. Browne, Phys. Rev. Lett. **104**, 030503 (2010).

[85] V. Veitch, S. A. H. Mousavian, D. Gottesman, and J. Emerson, New J. Phys. **16**, 013009 (2014).

[86] E. T. Campbell, H. Anwar, and D. E. Browne, Phys. Rev. X **2**, 041021 (2012).

[87] M. Howard and E. Campbell, Phys. Rev. Lett. **118**, 090501 (2017).

[88] S. F. E. Oliviero, L. Leone, A. Hamma, and S. Lloyd, npj Quantum Inf. **8**, 148 (2022).

[89] E. Knill and R. Laflamme, Phys. Rev. A **55**, 900 (1997).

[90] S. F. E. Oliviero, L. Leone, and A. Hamma, Phys. Rev. A **106**, 042426 (2022).

[91] L. Leone, S. F. E. Oliviero, and A. Hamma, arXiv:2204.02995.

[92] V. Strassen, Numer. Math. **13**, 354 (1969).

[93] J. Alman and V. V. Williams, arXiv:2010.05846.

[94] B. Collins, Int. Math. Res. Not. **2003**, 953 (2003).

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

[96] S. F. E. Oliviero, L. Leone, F. Caravelli, and A. Hamma, SciPost Phys. **10**, 076 (2021).