Loschmidt echo for probing operator hydrodynamics in heterogeneous structures

Algorithmiq

CONTENTS

I. INTRODUCTION

The steady progress in the hardware development has reached the phase, where the currently available digital quantum computers, though subject to noise, can potentially outperform classical simulation methods and provide scientifically valuable insights into the practically relevant problems that reduce to *an estimation of physical observables*. Given the noise constraints that restrict the problem size, the many-body and spin dynamics far from equilibrium is considered to be the most prominent candidate for leveraging the power of quantum computation [1, 2]. Nonetheless, some classical simulation methods still remain remarkably effective in predicting the dynamical observables in specific problems [3–9]. Taking into account also the problem-size rescaling approaches and the statistical arguments actual for the fully scrambling dynamics, a very narrow niche is left for a physical model to combine (i) practical relevance, (ii) high classical simulation complexity, (iii) technological feasibility to be implementable with the currently present quantum computers (see Fig. 1). The latter implies a requirement for a reasonably discernible output signal, so that the errors could be efficiently mitigated.

Semiscrambling quantum dynamics is a well suited candidate to reside in the emerging area for solving practically relevant problems with a quantum computer more efficiently than with the classical resources alone. This is due to a high degree of entanglement involved in the dynamics on one side and a retarded operator-support growth on the other side, which ensures the significantly large experimental signal in the presence of noise. Perfect conditions for

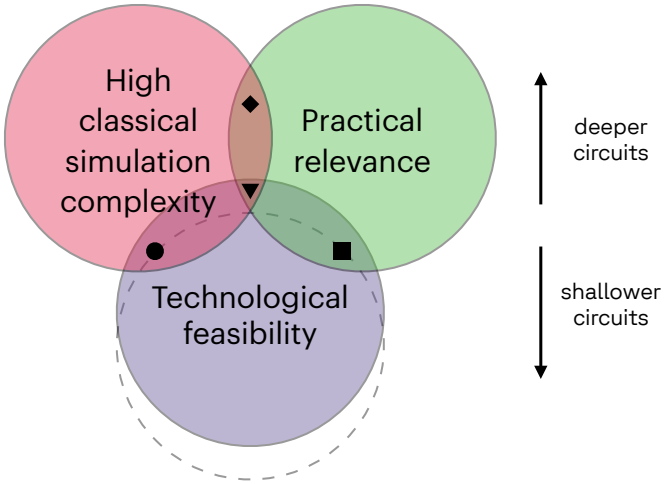

FIG. 1. **Quantum computing status.** Progress in the technology and quantum algorithms creates an emerging niche for solving practically relevant problems more efficiently than with the classical resources alone (▼). There are many problems of high complexity, e.g., the quantum phase estimation [10], that remain technologically infeasible (⧫). The previously reported experiments that were either classically hard (●, e.g., Refs. [11, 12]) or practically-oriented (■, e.g., Refs. [13, 14]) paved the way to the emergent research area with the both properties.

the observation of semiscrambling dynamics are present in the heterogeneous structures, where the interaction forms and strengths are distributed non-uniformly yet periodically akin to metamaterials [15] and Kitaev materials [16]. In this paper, we present the experimentally-friendly models for exploring the semiscrambling operator dynamics in heterogeneous structures compatible with the current quantum-computer architectures (see Sec. II).

The process of operator scrambling in heterogeneous structures is generally rather involved, with no direct evidence of what role particular sites and connections play in the operator spreading. We review the concept of the geometrical operator-support density and the operator hydrodynamics related to the operator norm—the only conserved quantity in a general unitary dynamics—in Sec. III. An important physical information about the main connections in the structure, through which the geometrical operator density primarily flows, could be highlighted in the interference experiments involving a scattering on the specific sites of the structure. The resulting time evolution of the operator contains scrambling (U_1) and unscrambling ($U_2^\dagger$) stages—manifesting in the destructive interference effects—that are hard to capture with the classical methods. On one hand, the tensor-network methods struggle due to the compression errors caused by the high entanglement generated in either of U_1 and $U_2^\dagger$. The tensor-network methods are also limited in the bond dimension in the attempt to accurately capture the combined evolution operator $U^\dagger = U_2^\dagger U_1$ in the from-the-middle-out contractions. The fast contraction method based on the belief propagation is approximate in non-tree tensor networks due to the loop correlations [17]. One the other hand, the Pauli-propagation-like methods [6–9] struggle due to the limited memory for tracking the operator branching and merging during the destructive interference.

This is a quantum-hardware experiment that enables to probe the eventual operator spreading under the dynamics $U^\dagger = U_2^\dagger U_1$ via a new probing technique that we refer to as the *operator Loschmidt echo* (OLE) explained in detail in Sec. IV. Briefly, the technique includes the evolution of the eigenvectors of a target operator O through the quantum circuit $UV_\delta U^\dagger$, where the generator G of the probe unitary $V_\delta = e^{-i\delta G}$ is associated with the property of the dynamical operator $U^\dagger OU$ under study. We show that the measurement of the target observable O at the end of such a δ-perturbed mirror circuit $UV_\delta U^\dagger$ enables one to extract the out-of-time-order correlation function (OTOC) $\propto \mathrm{tr}\{[G, U^\dagger OU]^\dagger [G, U^\dagger OU]\}$ as a coefficient for the signal decay with respect to the square of the perturbation strength, δ^2. The technique is similar in spirit though disparate in details from the ones exploited in Refs. [12, 14]: (i) the proposed technique allows measuring OTOC in the maximally mixed (infinite temperature) state not a single pure eigenstate of the operator O, (ii) the adjustable perturbation V_δ generally differs from a single Pauli gate, can be global, and ensures a stronger output signal for $\delta \ll 1$, with $\delta = 0$ serving as a perfect calibration point for any error mitigation. In the experiments conducted, the target operator O and the perturbation generator G have been chosen in such a way that they have geometrically disjoint Pauli supports—addressing the physics of how much of the operator-support density of the dynamical observable $U^\dagger OU$ lands on the probe support in the heterogeneous semiscrambling regime. In Sec. V, we outline the experimental design for measurement of the infinite-temperature OTOC in the dynamics $U^\dagger = U_2^\dagger U_1$.

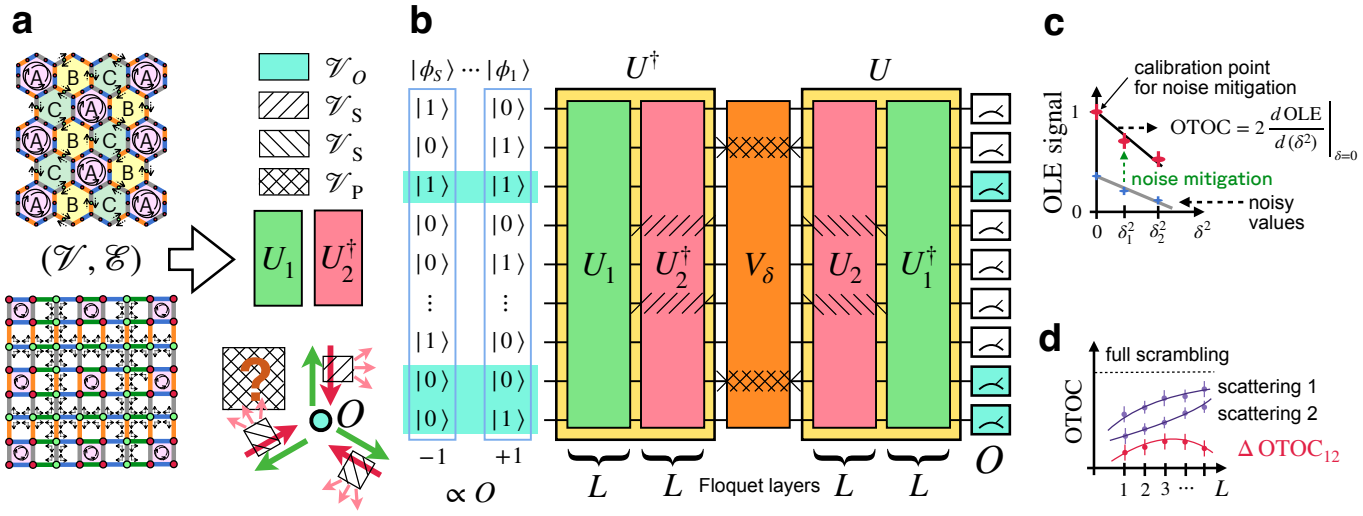

FIG. 2. **Problem of the operator hydrodynamics in materials and the proposed experimental design toward its solution. (a) Model.** Graph structure of the heterogeneous material defines the connectivity of each of L Floquet layers constituting scrambling (U_1) and unscrambling ($U_2^\dagger$) evolutions—manifesting in the destructive interference effects—that are hard to capture with the classical methods. The difference between U_1 and $U_2^\dagger$ in the overall dynamics $U^\dagger = U_2^\dagger U_1$ for an observable O (originally landing on vertices $\mathscr{V}_O$) is in the presence of scattering on some of the vertices ($\mathscr{V}_S$) or the edges ($\mathscr{E}_S$) in $U_2^\dagger$. The physical question is to probe how much of the operator density is located in the geometric area $\mathscr{V}_P$ (marked as $\boxed{?}$) as a result of the eventual operator spreading $U^\dagger O U$. The figure of merit is the out-of-time-order correlation function (OTOC) $\propto \mathrm{tr}\{[G, U^\dagger O U]^\dagger [G, U^\dagger O U]\}$ for the generator G acting nontrivially in the area $\mathscr{V}_P$. OTOC is probed with the operator Loschmidt echo (OLE) signal $\propto \mathrm{tr}[U V_\delta U^\dagger O U V_\delta^\dagger U^\dagger O]$ with $V_\delta = e^{-i\delta G}$. **(b) Proposed experimental setup for measuring the operator Loschmidt echo.** $O = Z^{\otimes \mathscr{V}_O}$. Preparation of a randomly chosen bitstring-state $|\phi_s\rangle \in \{|0\ldots0\rangle, \ldots, |1\ldots1\rangle\}$. Evolution of the state $|\phi_s\rangle$ under $UV_\delta U^\dagger$ results in $|\phi_s\rangle_{\mathrm{out}}$. Measurement of $|\phi_s\rangle_{\mathrm{out}}$ in the computational basis. Averaging the parity $\langle\phi_s|\,O\,|\phi_s\rangle \in \{+1, -1\}$ times $\langle\phi_s|_{\mathrm{out}}\,O\,|\phi_s\rangle_{\mathrm{out}}$ over S realizations provides the OLE estimation and its statistical error. **(c) Extraction of OTOC from OLE signals for different values of δ.** Global rescaling of the noisy signals with respect to that for $\delta = 0$ is the simplest noise mitigation technique. Other techniques have a calibration point at $\delta = 0$ since the whole circuit reduces to a mirror circuit. **(d) Revealing the physics of scattering.** Full scrambling assumption poses the upper bound for OTOC, $2|\mathscr{V}_P|$. Scattering on different elements of the structure (regimes 1 and 2) results in different OTOC values. The difference $\Delta\mathrm{OTOC}_{12}$ between those OTOCs in reasonably shallow circuits ($L \sim 5...10$) is challenging for *all* known classical methods.

Classical simulations of the OLE experiment can be divided into two classes: (i) addressing the output signal $\propto \mathrm{tr}[UV_\delta U^\dagger O U V_\delta^\dagger U^\dagger O]$ for the same choice of the perturbation strength δ as in the experiment, (ii) addressing the δ-independent OTOC value $\propto \mathrm{tr}\{[G, U^\dagger O U]^\dagger [G, U^\dagger O U]\}$, from which the output signal can be reconstructed. Due to the nature of the classical method, either (i) or (ii) is preferential from the viewpoint of efficiency. In Sec. VI, we review the belief-propagation tensor-network simulations in the Schrödinger picture suited for (i) as well as the Pauli propagation method, Monte Carlo single-Pauli-path method, and the method based on the purely statistical full-scrambling assumption that are all better suited for (ii). The list of other available classical simulation methods is much wider, with the Heisenberg-picture, the hybrid Schrödinger-and-Heisenberg-picture, and from-the-middle-out tensor-network simulations being some of them. Our preliminary analysis for an example heavy-hex topology model from Sec. II shows, however, that those tensor-network methods are not as efficient in terms of the convergence speed as the Schrödinger-picture simulations described in the paper.

In Sec. V, we present the circuit design for a proof-of-principle implementation of the OLE and the OTOC estimation in an operator dynamics on a heterogeneous heavy-hex structure, with the experiment being conducted by the IBM's team and the experiment details being available in Ref. [18]. In the case of the short-time evolution (shallow circuits), an agreement between the experimental results and the classical numerics is observed, whereas in the longer-time evolution (deeper circuits) a significant discrepancy is visible, with some of the classical predictions being beyond the physically meaningful bounds. In the case of the very deep circuits and lesser heterogeneity, we observe the tendency of the experimental results to approach the predictions obtained under the full-scrambling assumption, supporting the validity of the experimental results.

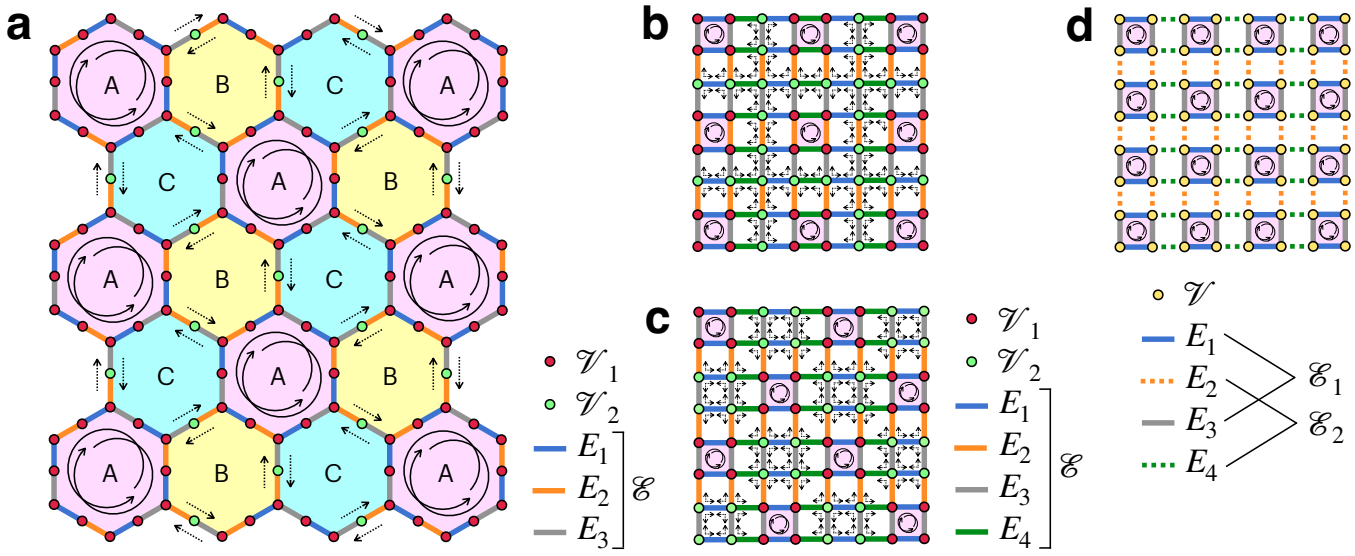

FIG. 3. **Heterogeneous structures.** Translationally invariant periodic lattices of inequivalent cells. Paradigmatic examples for heterogeneous operator dynamics are attained with **(a, b, c)** two sets of vertices or **(d)** two sets of edges resulting in the fast ($\mathscr{V}_1$, $\mathscr{E}_1$) or slow ($\mathscr{V}_2$, $\mathscr{E}_2$) propagation of the effective operator support through them. Operations in a single subset of edges E_j are commuting as they affect disjoint sets of vertices, making those operations implementable simultaneously and forming a single layer of interaction gates in quantum computing, where each vertex is associated with a qubit and the whole set of edges is compatible with qubits' connectivity implying no extra transpilation cost. **(a) A heavy-hex-lattice 2D material** compatible with the topology of IBM's quantum computers. **(b,c,d) Square-lattice 2D materials** compatible with the topology of, e.g., Google's, IQM's, and Rigetti's quantum computers. All the models are compatible with the reconfigurable qubits' connectivity enabled in trapped ion and neutral atom devices, such as Quantinuum's, QuEra's, and IonQ's quantum computers.

II. DYNAMICS IN THE STRUCTURED HETEROGENEOUS CIRCUITS

A. Graph representation

A general framework for the description and design of heterogeneous structures is based on the following graph representation. The graph vertices $\mathscr{V}$ correspond to the lattice sites, whereas the graph edges $\mathscr{E}$ encode the interaction pattern among the sites. In the digital-quantum-computer simulations of the spin-$\frac{1}{2}$ lattices, each graph vertex is associated with a qubit. The closer the pattern of graph edges $\mathscr{E}$ to the actual connectivity of two-qubit operations in the device, the shallower is the transpiled quantum circuit to be executed. In view of this, we focus on the graphs compatible with the topology of native gates in present quantum computers, some of them (heavy-hex and square ones) being depicted in Fig. 3. Should a quantum computer have a reconfigurable qubits' connectivity, a more extensive variability in the graph models is allowed (including but not limited to the 2D triangular and general 3D models). Availability of multiple-qubit native entangling gates would also extends the class of structures amenable to study.

Heterogeneity in the graph properties is attained through a non-uniform distribution of parameters in either the set of vertices $\mathscr{V}$, or the set of edges $\mathscr{E}$, or both sets. At the level of the computer native gates, this implies a heterogeneity in either the single-qubit gates ($\mathscr{V}$), or the two-qubit gates ($\mathscr{E}$), or both of them. In Sec. II B, we present several mechanisms of how the difference in vertex gates or edge gates implies different speeds of the effective-operator-support growth. The main idea is to separate the whole set of vertices $\mathscr{V}$ into *fast* (*slow*) information-transmission vertices $\mathscr{V}_1 \subset \mathscr{V}$ ($\mathscr{V}_2 \subset \mathscr{V}$) or to do a similar separation at the level of edges ($\mathscr{E} = \mathscr{E}_1 \cup \mathscr{E}_2$). All the unitary transformations are of the Floquet type and can be viewed as a kicked continuous-time evolution. Taking into account the digital-quantum-hardware implementation, we split the set of edges $\mathscr{E} = \cup_j E_j$ into disjoint sets corresponding to the commuting gates that are applied simultaneously and form a single layer in the quantum circuit.

The overall graph representation of the OLE experiment includes the following:

- The target observable O that is an operator acting non-trivially on a subset of qubits ($\mathscr{V}_O$). Typically, O is a Pauli operator (see Sec. IV C), e.g., $O = \bigotimes_{q\in\mathscr{V}_O} Z_q$ or $O = \bigotimes_{q\in\mathscr{V}_O} X_q$.

- A Floquet layer $U_{\rm FL}$ as a part of the unitary dynamics U. Typically, $U_{\rm FL}=\prod_{E\subset\mathscr{E}}\big(\bigotimes_{(q_1,q_2)\in E}u_{q_1q_2}\big)\big(\bigotimes_{q\in\mathscr{V}}u_q\big)$ or $U_{\rm FL}=\prod_{E\subset\mathscr{E}}\big(\bigotimes_{(q_1,q_2)\in E}u_{q_1q_2}u_{q_1}u_{q_2}\big)$, where $u_{q_1q_2}$ is a native two-qubit gate acting on qubits q_1 and q_2 and u_q is a single-qubit gate.

- The difference between the scrambling (U_1) and unscrambling ($U_2^\dagger$) parts of the dynamics $U^\dagger=U_2^\dagger U_1$ is in the presence of scattering on some of the vertices or the edges in $U_2^\dagger$. If $U_{\rm FL}$ is a Floquet layer in U_1, then its scattered version $U_{\rm FLS}$ in U_2 reads $U_{\rm FLS}=\big(\bigotimes_{q\in\mathscr{V}_{\rm S}}v_q\big)U_{\rm FL}\big(\bigotimes_{q\in\mathscr{V}_{\rm S}}v_q^\dagger\big)$ or $U_{\rm FLS}=\big(\bigotimes_{(q_1,q_2)\in\mathscr{E}_{\rm S}}v_{q_1q_2}\big)U_{\rm FL}\big(\bigotimes_{(q_1,q_2)\in\mathscr{E}_{\rm S}}v_{q_1q_2}^\dagger\big)$ for scatterings on the set of vertices $\mathscr{V}_{\rm S}$ or a set of the connections $\mathscr{E}_{\rm S}$, respectively. Typically, the structure of $U_{\rm FLS}$ is similar to that of $U_{\rm FL}$, with the difference being in some of the qubit-local parameters only (where the scattering takes place geometrically). The effect of scattering in the following sections boils down to a sign change for some of the parameters in $U_{\rm FL}$.

- Perturbation $V_\delta=e^{-i\delta G}$ is generated by a Hermitian operator G acting non-trivially on a subset of qubits ($\mathscr{V}_{\rm P}$). Typically, $G=\sum_{q\in\mathscr{V}_{\rm P}}Z_q$ or $G=\sum_{q\in\mathscr{V}_{\rm P}}X_q$ implying $V_\delta=\bigotimes_{q\in\mathscr{V}_{\rm P}}e^{-i\delta Z_q}$ or $V_\delta=\bigotimes_{q\in\mathscr{V}_{\rm P}}e^{-i\delta X_q}$, respectively.

To resume, the graph representation of the gate-based model consists of the vertices $\mathscr{V}$, the edges $\mathscr{E}$, the observable localization $\mathscr{V}_O$, the scattering localization $\mathscr{V}_{\rm S}$ or $\mathscr{E}_{\rm S}$, and the perturbation localization $\mathscr{V}_{\rm P}$. The problem heterogeneity is encoded in either the fast and slow vertices $\mathscr{V}_1,\mathscr{V}_2\subset\mathscr{V}$, or the fast and slow edges $\mathscr{E}_1,\mathscr{E}_2\subset\mathscr{E}$, or a combination of two. Given the digital nature of quantum gates, the set of edges $\mathscr{E}$ is convenient to split into the subsets of simultaneously implemented gates, $\mathscr{E}=\cup_j E_j$.

B. Gate-based models

In the following sections, we define the Floquet layer structure in terms of the typical native entangling gates exploited in the current quantum computers. Once the Floquet unitary $U_{\rm FL}$ and its scattering version $U_{\rm FLS}$ are determined, the dynamics U_1 (U_2) consists of L repeatedly applied Floquet layers $U_{\rm FL}$ ($U_{\rm FLS}$),

$$U_1(L)=(U_{\rm FL})^L,\qquad U_2(L)=(U_{\rm FLS})^L,\qquad U^\dagger(2L)=U_2(L)^\dagger U_1(L)=(U_{\rm FLS}^\dagger)^L(U_{\rm FL})^L. \tag{1}$$

In general, the number of Floquet layers does not have to be the same in U_1 and U_2, but hereafter we restrict ourselves to that scenario.

1. *Models with the continuously parameterized R_{ZZ}-like and Mølmer-Sørensen-like native entangling gates*

The Mølmer-Sørensen gate applied to a pair of ion qubits is essentially the $R_{XX}(\theta)$ gate, which is equivalent to the $R_{ZZ}(\theta)$ gate up to the single qubit rotations. Therefore, the above entangling native gates could be routinely used to implement the Floquet layer

$$U_{\rm FL}^{\rm ZZX}(J_1,J_2,h,b_1,b_2)=\prod_{E\subset\mathscr{E}}\bigotimes_{(q_1,q_2)\in E}e^{-iJ_{(q_1,q_2)}Z_{q_1}Z_{q_2}}e^{-ih(Z_{q_1}+Z_{q_2})}e^{-i(b_{q_1}X_{q_1}+b_{q_2}X_{q_2})}, \tag{2}$$

$$J_{(q_1,q_2)}=\begin{cases}J_1, & (q_1,q_2)\in\mathscr{E}_1,\\ J_2, & (q_1,q_2)\in\mathscr{E}_2,\end{cases}\qquad b_q=\begin{cases}b_1, & q\in\mathscr{V}_1,\\ b_2, & q\in\mathscr{V}_2,\end{cases} \tag{3}$$

where the heterogeneity can be attained either at the level of edges when the parameter $J_{(q_1,q_2)}$ varies across the subsets of edges, or at the level of vertices when the X-field depends on the qubit position, or via a combination of the two. The following result allows to navigate rationally through the parameter landscape: it clarifies that the information propagation through the qubit q is blocked if $b_q=0$ or $b_q=\frac{\pi}{2}$.

Proposition 1. *Let U be a unitary operator corresponding to a quantum circuit on a graph $(\mathscr{V},\mathscr{E})$ with the gates $R_{ZZ}(\theta)=e^{-i\theta ZZ/2}$, $R_Z(\zeta)=e^{-i\zeta Z/2}$, and $R_X(\varphi)=e^{-i\varphi X/2}$, where $\theta,\zeta,\varphi\in\mathbb{R}$. Let O be an operator that acts nontrivially only on vertices $\mathscr{V}_O\subset\mathscr{V}$, then the operator $U^\dagger OU$ acts trivially on all qubits $q\in\mathscr{V}\backslash\mathscr{V}_O$ such that every graph path $(q,\ldots,q')\subset\mathscr{E}$ connecting q and $q'\in\mathscr{V}_O$ contains a qubit q_* such that $\varphi=\pi n$, $n\in\mathbb{Z}$ for every $R_X(\varphi)$ gate acting on qubit q_*.*

Proof. Consider two consecutive edges (q_1,q_*) and (q_*,q_2) on a graph path $(q,\ldots,q')\subset\mathscr{E}$ connecting q and $q'\in\mathscr{V}_O$. Since the sequential application of circuit gates to the operator O generally results in the growth of its causal cone, suppose that after some number of gate application the evolved operator O_* acts trivially (as I) on q_1 and q_* and

notrivially (as X, or Y, or Z) on q_2. The application of any gates on q_1 and q_* does not change the Pauli operator support of O_*, and neither does the application of local one-qubit gates on q_2. The application of the entangling $R_{ZZ}(\theta)$ gate to qubits q_* and q_2 can only result in the non-trivial action of the evolved operator $R^\dagger_{Z_{q_*}Z_{q_2}}(\theta)O_*R_{Z_{q_*}Z_{q_2}}(\theta)$ on qubit q_* in the form of the Z operator. Indeed, $R^\dagger_{ZZ}(\theta)\cdot IX\cdot R_{ZZ}(\theta)=\cos\theta\cdot IX+\sin\theta\cdot ZY$, $R^\dagger_{ZZ}(\theta)\cdot IY\cdot R_{ZZ}(\theta)=\cos\theta\cdot IY-\sin\theta\cdot ZX$, and $R^\dagger_{ZZ}(\theta)\cdot IZ\cdot R_{ZZ}(\theta)=IZ$. The Z operator on qubit q_* commutes with the gates $R_Z(\zeta)$ and $R_{ZZ}(\theta)$ affecting it, whereas every $R_X(\varphi)$ gate acting on qubit q_* is (up to a phase) either I (if $\varphi=2\pi n$, $n\in\mathbb{Z}$) or X (if $\varphi=\pi+2\pi n$, $n\in\mathbb{Z}$), and neither of them changes the Z-action of the operator on qubit q_* ($XZX=-Z$). Therefore, the only possible nontrivial component of the dynamical observable on qubit q_* is Z, which cannot propagate along the edge (q_1,q_*) to q_1 and, consequently, to q along the path $(q,\dots,q_1)\subset(q,\dots,q')$ unless there are alternative information propagation paths. □

According to Proposition 1, the causal cone of the observable stops expanding once the qubit-q local X-field b_q equals 0 or a multiple of $\frac{\pi}{2}$. Should $b_q\ll 1$, the information is able to *leak* through that qubit in the retarded way as the Z operator on qubit q is transformed by e^{-ibX} into the component $\sin(2b)Y$, and the operator IY in turn is transformed by e^{-iJZZ} into an operator containing the propagated term $\sin(2J)ZX$. Therefore, the operator-leakage probability through a connection (q_1,q_2) in one Floquet layer can be roughly estimated as $\sin^2(2b_{q_1})\sin^2(2b_{q_2})\sin^2(2J_{(q_1,q_2)})$. In order to obtain a heterogeneous dynamics with the slow and fast information-propagation channels, the choice of parameters (3) is to ensure an imbalance in the operator-leakage probabilities. Unless otherwise stated, we assume $\frac{\pi}{4}\geq|b_{q\in\mathscr{V}_1}|>|b_{q\in\mathscr{V}_2}|>0$ if J-parameters are distributed uniformly, $\frac{\pi}{4}\geq|J_{(q_1,q_2)\in\mathscr{E}_1}|>|J_{(q_1,q_2)\in\mathscr{E}_2}|>0$ if b-parameters are distributed uniformly.

Locating Z-scattering on the (fast) vertices $\mathscr{V}_1$ or (slow) vertices $\mathscr{V}_2$, we get different physical regimes

$$U^{\mathrm{ZZX}}_{\mathrm{FLS}}(J_1,J_2,h,b_1,b_2)=\begin{cases}U^{\mathrm{ZZX}}_{\mathrm{FL}}(J_1,J_2,h,-b_1,b_2), & \mathscr{V}_{\mathrm{S}}=\mathscr{V}_1,\\ U^{\mathrm{ZZX}}_{\mathrm{FL}}(J_1,J_2,h,b_1,-b_2), & \mathscr{V}_{\mathrm{S}}=\mathscr{V}_2.\end{cases}\tag{4}$$

Eqs. (2), (3), (4) define the dynamics (1) for different physical regimes, with the model heterogeneity and the position of scattering vertices affecting the resulting distribution of the operator-support density for $U^\dagger OU$. The latter is probed by the operator-Loschmidt-echo circuit $UV_\delta U^\dagger$, where the perturbation V_δ acts nontrivially on the vertices $\mathscr{V}_{\mathrm{P}}$ that are typically disjoint from the operator vertices $\mathscr{V}_O$ in order to probe and compare the physics of operator hydrodynamics in the cases of scattering on different elements of the circuit.

Note that some entangling layers in U cancel at the seam between $U_2^\dagger$ and U_1, facilitating tensor-network simulations. However, the following XZZX symmetric arrangement of the gates (akin to a higher-order-Trotter decomposition) reduces the number of canceling gates (down to 0 in some topologies) making tensor-network simulations harder:

$$U^{\mathrm{XZZX}}_{\mathrm{FL}}(J_1,J_2,h,b_1,b_2)=\prod_{E\subset\mathscr{E}}\bigotimes_{(q_1,q_2)\in E}e^{-i(b_{q_1}X_{q_2}+b_{q_2}X_{q_2})/2}e^{-iJ_{(q_1,q_2)}Z_{q_1}Z_{q_2}}e^{-ih(Z_{q_1}+Z_{q_2})}e^{-i(b_{q_1}X_{q_1}+b_{q_2}X_{q_2})/2},\tag{5}$$

$$J_{(q_1,q_2)}=\begin{cases}J_1, & (q_1,q_2)\in\mathscr{E}_1,\\ J_2, & (q_1,q_2)\in\mathscr{E}_2,\end{cases}\qquad b_q=\begin{cases}b_1, & q\in\mathscr{V}_1,\\ b_2, & q\in\mathscr{V}_2,\end{cases}\tag{6}$$

$$U^{\mathrm{XZZX}}_{\mathrm{FLS}}(J_1,J_2,h,b_1,b_2)=\begin{cases}U^{\mathrm{XZZX}}_{\mathrm{FL}}(J_1,J_2,h,-b_1,b_2), & \mathscr{V}_{\mathrm{S}}=\mathscr{V}_1,\\ U^{\mathrm{XZZX}}_{\mathrm{FL}}(J_1,J_2,h,b_1,-b_2), & \mathscr{V}_{\mathrm{S}}=\mathscr{V}_2.\end{cases}\tag{7}$$

2. *Models with the CZ-like native entangling gates*

Since the CZ gate is equivalent to the CX and the echoed cross-resonance gate up to local single-qubit rotations, all those native gates can be considered within a single class of entangling native gates. The gate $R_{ZZ}(\frac{\pi}{2})=e^{-i\frac{\pi}{4}ZZ}=e^{-i\frac{\pi}{4}}\cdot S\otimes S\cdot CZ$ belongs to the same class too. Therefore, all these native gates could be used to implement the Floquet layer

$$\begin{aligned}U^{\mathrm{CZX}}_{\mathrm{FL}}(h,b_1,b_2) &= U^{\mathrm{ZZX}}_{\mathrm{FL}}(J_1=\tfrac{\pi}{4},J_2=\tfrac{\pi}{4},h,b_1,b_2)\\ &= \prod_{E\subset\mathscr{E}}\bigotimes_{(q_1,q_2)\in E}e^{-i\frac{\pi}{4}Z_{q_1}Z_{q_2}}e^{-ih(Z_{q_1}+Z_{q_2})}e^{-i(b_{q_1}X_{q_2}+b_{q_2}X_{q_2})},\qquad b_q=\begin{cases}b_1, & q\in\mathscr{V}_1,\\ b_2, & q\in\mathscr{V}_2,\end{cases}\end{aligned}\tag{8}$$

where the heterogeneity is due to the X-field parameter b_q dependent on the qubit position q. This model is a particular case of one considered in the previous section, so Proposition 1 is applicable to it too. Therefore, the operator leakage through a vertex q is controlled by the b_q-parameter. Unless otherwise stated, we assume $\frac{\pi}{4}\geq|b_{q\in\mathscr{V}_1}|>|b_{q\in\mathscr{V}_2}|>0$. The smaller $|b_q|$ the slower the leakage.

Locating Z-scattering on the (fast) vertices $\mathscr{V}_1$ or (slow) vertices $\mathscr{V}_2$, we get different physical regimes

$$U^{\mathrm{CZX}}_{\mathrm{FLS}}(h,b_1,b_2)=\begin{cases} U^{\mathrm{CZX}}_{\mathrm{FL}}(h,-b_1,b_2), & \mathscr{V}_{\mathrm{S}}=\mathscr{V}_1,\\ U^{\mathrm{CZX}}_{\mathrm{FL}}(h,b_1,-b_2), & \mathscr{V}_{\mathrm{S}}=\mathscr{V}_2.\end{cases} \tag{9}$$

Eqs. (8) and (9) define the dynamics (1) for different physical regimes, with the model heterogeneity and the position of scattering vertices affecting the resulting distribution of the operator-support density for $U^\dagger OU$. The latter is probed by the operator-Loschmidt-echo circuit $UV_\delta U^\dagger$, where the perturbation V_δ acts nontrivially on the vertices $\mathscr{V}_{\mathrm{P}}$ that are typically disjoint from the operator vertices $\mathscr{V}_O$ in order to probe and compare the physics of operator hydrodynamics in the cases of scattering on different elements of the circuit.

Note that some entangling layers in U cancel at the seam between $U_2^\dagger$ and U_1, facilitating tensor-network simulations. However, the following XCZX symmetric arrangement of the gates (akin to a higher-order-Trotter decomposition) reduces the number of canceling gates (down to 0 in some topologies) making tensor-network simulations harder:

$$U^{\mathrm{XCZX}}_{\mathrm{FL}}(h,b_1,b_2)=\prod_{E\subset\mathscr{E}}\bigotimes_{(q_1,q_2)\in E} e^{-i(b_{q_1}X_{q_2}+b_{q_2}X_{q_2})/2}e^{-i\frac{\pi}{4}Z_{q_1}Z_{q_2}}e^{-ih(Z_{q_1}+Z_{q_2})}e^{-i(b_{q_1}X_{q_1}+b_{q_2}X_{q_2})/2}, \tag{10}$$

$$U^{\mathrm{XCZX}}_{\mathrm{FLS}}(h,b_1,b_2)=\begin{cases} U^{\mathrm{XCZX}}_{\mathrm{FL}}(h,-b_1,b_2), & \mathscr{V}_{\mathrm{S}}=\mathscr{V}_1,\\ U^{\mathrm{XCZX}}_{\mathrm{FL}}(h,b_1,-b_2), & \mathscr{V}_{\mathrm{S}}=\mathscr{V}_2,\end{cases}\qquad b_q=\begin{cases} b_1, & q\in\mathscr{V}_1,\\ b_2, & q\in\mathscr{V}_2.\end{cases} \tag{11}$$

3. *Models with the R_{XX+YY} and iSWAP-like native entangling gates*

Consider the gate $R_{XX+YY}(\theta)=e^{-i\theta(XX+YY)/2}$, which reduces to the iSWAP gate (up to a global phase) if $\theta=-\frac{\pi}{2}$. Two $R_{XX+YY}(\theta)$ gates acting on the overlapping qubit pairs (q_1,q_2) and (q_2,q_3) do not commute whenever the gates are entangling, i.e., $\theta\neq\pi n$, $n\in\mathbb{Z}$. For this reason, the simplest heterogeneous dynamics is established through a nonuniform distribution of angles θ, namely,

$$U^{\mathrm{XX+YY|Z}}_{\mathrm{FL}}(J_1,J_2,h_1,h_2)=\prod_{E\subset\mathscr{E}}\bigotimes_{(q_1,q_2)\in E} e^{-iJ_{(q_1,q_2)}(X_{q_1}X_{q_2}+Y_{q_1}Y_{q_2})}e^{-i(h_{q_1}Z_{q_1}+h_{q_2}Z_{q_2})},\qquad J_{(q_1,q_2)}=\begin{cases} J_1, & (q_1,q_2)\in\mathscr{E}_1,\\ J_2, & (q_1,q_2)\in\mathscr{E}_2.\end{cases} \tag{12}$$

The speed of information propagation through edges $\mathscr{E}_1$ is faster than that through edges $\mathscr{E}_2$ if $\frac{\pi}{4}\geq|J_1|>|J_2|\geq 0$. Apparently, the operator spreading stops through the edges $\mathscr{E}_2$ if $J_2=0$. If $h_1=h_2$, then the dynamics preserves the total magnetization $\sum_q Z_q$. The effective scattering on the edges $\mathscr{E}_1$ or $\mathscr{E}_2$ corresponds to

$$U^{\mathrm{XX+YY|Z}}_{\mathrm{FLS}}(J_1,J_2,h_1,h_2)=\begin{cases} U^{\mathrm{XX+YY|Z}}_{\mathrm{FL}}(J_1,J_2,-h_1,h_2), & \mathscr{E}_{\mathrm{S}}=\mathscr{E}_1,\\ U^{\mathrm{XX+YY|Z}}_{\mathrm{FL}}(J_1,J_2,h_1,-h_2), & \mathscr{E}_{\mathrm{S}}=\mathscr{E}_2,\end{cases} \tag{13}$$

with the scattering operator being Z on a single qubit from the edge $(q_1,q_2)\in\mathscr{E}_{\mathrm{S}}$ (should a topology be compatible with that).

Should the iSWAP gate be the native gate of the device, then the following intricate dynamics exhibits the adjustable heterogeneity:

$$U^{\mathrm{iSWAP|\Delta Z}}_{\mathrm{FL}}(h,\Delta h_1,\Delta h_2)=\prod_{E\subset\mathscr{E}}\bigotimes_{(q_1,q_2)\in E} e^{-i\frac{\pi}{2}(X_{q_1}X_{q_2}+Y_{q_1}Y_{q_2})}e^{-ih(Z_{q_1}+Z_{q_2})-i\Delta h_{(q_1,q_2)}(Z_{q_1}-Z_{q_2})}e^{i\frac{\pi}{2}(X_{q_1}X_{q_2}+Y_{q_1}Y_{q_2})}, \tag{14}$$

$$\Delta h_{(q_1,q_2)}=\begin{cases} \Delta h_1, & (q_1,q_2)\in\mathscr{E}_1,\\ \Delta h_2, & (q_1,q_2)\in\mathscr{E}_2.\end{cases} \tag{15}$$

The information propagation through the qubit pair (q_1,q_2) is blocked if $\Delta h_{(q_1,q_2)}=0$ because no entanglement is created across the link in this case. The larger $|\Delta h_{(q_1,q_2)}|\in[0,\frac{\pi}{2}]$ the more intense the operator leakage. The importance of different links in the operator spreading can be highlighted in the process

$$U^{\mathrm{iSWAP|\Delta Z}}_{\mathrm{FLS}}(h,\Delta h_1,\Delta h_2)=\begin{cases} U^{\mathrm{iSWAP|\Delta Z}}_{\mathrm{FL}}(h,-\Delta h_1,\Delta h_2), & \mathscr{E}_{\mathrm{S}}=\mathscr{E}_1,\\ U^{\mathrm{iSWAP|\Delta Z}}_{\mathrm{FL}}(h,\Delta h_1,-\Delta h_2), & \mathscr{E}_{\mathrm{S}}=\mathscr{E}_2.\end{cases} \tag{16}$$

C. Probing operator spreading with local and global OTOCs

Should G be a single qubit operator, for instance $G = Z_q$ for some qubit position q, then OTOC would be sensitive to a wavefront of the Heisenberg-picture operator spreading $O_0(t) = U_0^\dagger(t) O U_0(t)$ from the region qubits $\mathcal{Q}$ where O acts nontrivially. The OTOC is zero beyond the observable light cone and is bounded from above by 4 if $\|O\|_2^2 = 2^N$ (as takes place, e.g., for $O = Z_\mathcal{Q}$). The profile of local OTOC along the grid of qubits can be rather nonuniform even in the case of local interactions constituting $U_0(t)$ if the interactions between qubits are heterogeneous in space. Heterogeneous systems can exhibit a regime of semi-scrambling, in which some regions of qubits induce fast chaotic-like scrambling, whereas others have much slower diffusive effect.

In case of the global generator G, for instance $G = \sum_{q\in\mathcal{A}} Z_q$ that acts nontrivially in some region $\mathcal{A}$ of qubits, the corresponding global OTOC can be roughly estimated as the area under the local OTOC in this region [19].

III. HYDRODYNAMICS OF THE OPERATOR-SUPPORT DENSITY

Keeping in mind some connectivity graph for qubits on an actual hardware, e.g., the heavy hexagonal topology, let $\mathbf{r}$ denote a geometrical position of the qubit in the given topology. In the case of 1D line, $\mathbf{r}$ is merely a position x on the line, i.e., $x \in \{0, \ldots, N-1\}$ for an N-qubit register. In the case of some 2D topology, $\mathbf{r} = (x, y)$. Similarly, in 3D, $\mathbf{r} = (x, y, z)$. In general, for a finite number of qubits, N, we have a set of available qubit positions $\mathcal{R}$ such that $|\mathcal{R}| = N$.

Associated with the qubit register is the space $\mathbb{C}^{2^N\times 2^N}$ of operators acting on N-qubit states. The Pauli operators $P \equiv \bigotimes_{\mathbf{r}\in\mathcal{R}} P_\mathbf{r}$, where $P_\mathbf{r} = I, X, Y, Z$, form an orthogonal basis in the operator space with respect to the Hilbert–Schmidt scalar product, namely, $\langle P|P'\rangle = 2^N \delta_{PP'}$. Any operator O admits a decomposition

$$O = \sum_P c_P P \tag{17}$$

so that $2^N \sum_P |c_P|^2 = \|O\|_2^2$.

Let $U(t)$ be a unitary evolution operator for time $t \in \mathbb{R}_+$. The operator dynamics in the Heisenberg picture reads

$$O(t) \equiv U^\dagger(t) O U(t) = \sum_P c_P(t) P, \tag{18}$$

where $\sum_P |c_P(t)|^2 = 2^{-N}\|O\|_2^2$ regardless of time t. If the initial operator O is a single Pauli string, then $\|O\|_2^2 = 2^N$ and $\sum_P |c_P(t)|^2 = 1$ throught the evolution. Therefore, $|c_P(t)|^2$ is a time-dependent discrete probability distribution in the 4^N-dimensional space of Pauli strings P.

The (geometrical) Pauli support for a Pauli string P is defined as a set of qubit positions where the operator P acts nontrivially, i.e., $\mathrm{supp}(P) = \{\mathbf{r} \in \mathcal{R} | P_\mathbf{r} \neq I\}$. By $w(P)$ denote the Pauli weight of P, i.e., $w(P) = \sum_{\mathbf{r}: P_\mathbf{r}\neq I} 1$. Clearly, $w(P) = |\mathrm{supp}(P)|$. The operator-support density ρ for a single Pauli operator P is defined as $\rho_P(\mathbf{r}) \equiv \frac{1}{|\mathrm{supp}(P)|} \times \begin{cases} 1, & \mathbf{r} \in \mathrm{supp}(P), \\ 0, & \mathbf{r} \notin \mathrm{supp}(P), \end{cases}$ and shows that the non-trivial action of the Pauli string is spread over the particular subset of qubits. Note that $\sum_\mathbf{r} \rho_P(\mathbf{r}) = 1$ irrespective of P.

The operator-support density for a general time-dependent operator $O(t)$ in Eq. (18) is defined through

$$\rho(\mathbf{r}, t) = \sum_P |c_P(t)|^2 \rho_P(\mathbf{r}) = \sum_{P:\ \mathbf{r}\in\mathrm{supp}(P)} |c_P(t)|^2 |\mathrm{supp}(P)|^{-1}. \tag{19}$$

For a fixed time moment t, Eq. (19) defines a legit probability distribution since $\rho(\mathbf{r}, t) \geq 0$ and $\sum_\mathbf{r} \rho(\mathbf{r}, t) = \sum_P |c_P(t)|^2 \sum_\mathbf{r} \rho_P(\mathbf{r}) = 1$, where we have used the properties $\sum_\mathbf{r} \rho_P(\mathbf{r}) = 1$ and $\sum_P |c_P(t)|^2 = 1$.

Consider an operator dynamics with a local observable O at time $t = 0$. In the process of evolution, the operator support typically spreads as interactions tend to delocalize. The continuity equation for the operator-support density,

$$\frac{\partial \rho}{\partial t} + \mathrm{div}\, \mathbf{j} = 0, \tag{20}$$

defines the density current, $\mathbf{j}(\mathbf{r}, t)$.

Let $S_\mathrm{Sh}(\varrho)$ be the Shannon entropy of the density ϱ. Then an *effective* area of the operator-support density can be defined as $e^{S_\mathrm{Sh}(\varrho)}$. The *effective* (geometrical) Pauli support of the operator $O(t)$ consists of such points $\mathbf{r}$, where $\varrho(\mathbf{r}, t) \geq e^{-S_\mathrm{Sh}(\varrho)}$. The change in the effective Pauli support can be probed with the operator-Loschmidt-echo experiment.

IV. LOSCHMIDT ECHO

This section is solely dedicated to probing the operator spreading in the echo-type experiments, with no regard to the underlying time evolution $U = U(t)$ that governs the operator dynamics at the first place. It is Sec. V, where we combine the measurement of global OTOC with the models $U = U_2^\dagger U_1$ considered in Sec. II.

A. Loschmidt echo for states

The return probability $\left|\langle\psi| U_\delta^\dagger(t)U_0(t)|\psi\rangle\right|^2$ of a pure state $|\psi\rangle$ under perturbed time-reversed unitary evolution $U_\delta^\dagger(t)U_0(t)$ is the Loschmidt echo signal [20–22] that exhibits high sensitivity to the perturbation strength δ and, therefore, has been proposed as a tool in identifying chaotic dynamics, probing dynamical phase transitions, and enhancing quantum metrology. However, the echo signal in many-body systems typically decays exponentially in the system size, thus hindering its experimental observation unless scanning over smaller subsystems is done [23]. Implementation of the conventional Loschmidt echo for states on an ideal quantum computer would imply repeated preparation of the multiqubit state $|0\rangle^{\otimes N}$, its evolution in the gate-based manner according to the time-reversed unitary operator $U_\delta^\dagger(t)U_0(t)$, and measurement of the resulting state in the computational basis $\{0,1\}^N$ with the aim to quantify the frequency $f_{\rm LES}$ of events in which the all-zero bitstring $0^{\times N}$ is observed as a measurement outcome, $f_{\rm LES} \sim e^{-\delta N t}$. The current generation of quantum computers struggles with conducting such an experiment because of the prohibitive requirements with respect to the initialization- and measurement-repetition rates as well as the gate and readout fidelities.

B. Operator Loschmidt echo (OLE)

In contrast to the return probability for states, the operator Loschmidt echo (OLE) quantifies the normalised overlap (the Hilbert-Schmidt scalar product) of the original operator O with its perturbed time-reversed version,

$$L = \frac{\mathrm{tr}\left(O^\dagger \times U_\delta(t)U_0^\dagger(t)OU_0(t)U_\delta^\dagger(t)\right)}{\mathrm{tr}\left(O^\dagger O\right)}. \tag{21}$$

OLE can also be seen as the overlap between the Heisenberg-picture observable $O_0(t) = U_0^\dagger(t)OU_0(t)$ and its perturbed counterpart $O_\delta(t) = U_\delta^\dagger(t)OU_\delta(t)$, namely, $L = \mathrm{tr}\big(O_0(t)O_\delta^\dagger(t)\big)/\mathrm{tr}(O^\dagger O)$. The denominator of Eq. (21) takes into account that the Schatten 2-norm (the Frobenius norm) is preserved by the unitary dynamics, i.e., $\|O\|_2 = \|O_0(t)\|_2 = \|O_\delta(t)\|_2 = \sqrt{\mathrm{tr}\left(O^\dagger O\right)}$. Upon vectorisation of the $2^N \times 2^N$-dimensional operator space, Eq. (21) resembles the return amplitude in the echo signal for 4^N-dimensional "states", $L = \langle\langle O_\delta(t)|O_0(t)\rangle\rangle/\langle\langle O|O\rangle\rangle$, though it is drastically disparate from the Loschmidt echo for states from the viewpoint of experimental feasibility. The novelty of our approach is in the fact that we avoid dealing with N ququarts (4-dimensional systems), which would be mathematically equivalent to the Loschmidt echo for $2N$-qubit pure states, but rather implement the measurement procedure for Eq. (21) at the level of N qubits by exploiting the ensemble representation of mixed states and conventional measurements.

C. Measurement of OLE

Let the operator O be Hermitian traceless operator with eigenvalues ±1, for instance, a Pauli string on N qubits. Then $\|O\|_2^2 = \mathrm{tr}(O^2) = 2^N$ and Eq. (21) reduces to

$$L = \frac{1}{2^N}\mathrm{tr}\left(O_0(t)O_\delta(t)\right). \tag{22}$$

Let $\mathcal{V}^\pm := \{|\psi_i^\pm\rangle\}_i$ be a set of orthonormal eigenvectors of O corresponding to the eigenvalue ±1. Then the cardinality $|\mathcal{V}^\pm| = 2^{N-1}$, and

$$\varrho_{\pm O} = \frac{I^{\otimes N} \pm O}{2^N} = \frac{1}{2^{N-1}}\sum_i |\psi_i^\pm\rangle\langle\psi_i^\pm| \tag{23}$$

are genuine density operators because they both have unit trace ($\mathrm{tr}(\varrho_{\pm O}) = 1$) and nonnegative eigenvalues 0 and $2^{-(N-1)}$ each of degeneracy 2^{N-1} ($\varrho_{\pm O} \geqslant 0$). Despite the fact that the state (23) is generally mixed, its preparation on quantum hardware is straightforward thanks to the ensemble representation in terms of the eigenvectors of O.

Similarly, one can effectively prepare on a quantum computer the operator O itself by sampling uniformly S eigenvectors $|\phi_s\rangle \in \mathcal{V}^+ \cup \mathcal{V}^-$ and assigning the corresponding eigenvalues $\langle\phi_s| O |\phi_s\rangle = \pm 1$ to them, namely,

$$\frac{1}{S}\sum_{s=1}^{S} \underbrace{\langle\phi_s| O |\phi_s\rangle}_{\pm 1} |\phi_s\rangle \langle\phi_s| \longrightarrow \frac{1}{2^N} O \quad \text{if} \quad S \longrightarrow \infty. \tag{24}$$

There is no extra sampling cost for the implementation of Eq. (24) even though roughly half of the coefficients is negative. This is because $\frac{1}{2^N} O = \frac{1}{2}(\varrho_{+O} - \varrho_{-O})$ is a weighted difference of two density operators, with weights summing to unity.

Let $|\phi_s(t)\rangle := U_\delta(t) U_0^\dagger(t) |\phi_s\rangle$ denote the time-reversed perturbed evolution of eigenstates $|\phi_s\rangle$ of O. Substituting the representation (24) for O in $O_0(t) = U_0^\dagger(t) O U_0(t)$ from Eq. (22), we obtain the operator Loschmidt echo

$$L = \lim_{S\to\infty} L_S \tag{25}$$

through its estimator

$$L_S = \frac{1}{S}\sum_{s=1}^{S} \underbrace{\langle\phi_s| O |\phi_s\rangle}_{\pm 1} \times \langle\phi_s(t)| O |\phi_s(t)\rangle \,. \tag{26}$$

The OLE estimator L_S is experimentally accessible by means of measuring O in the evolved states $|\phi_s(t)\rangle$, and its estimation error quickly decreases as $\propto 1/\sqrt{S}$. It is the correlation between $\langle\phi_s| O |\phi_s\rangle$ and $\langle\phi_s(t)| O |\phi_s(t)\rangle$ that constitutes the echo experiment and should be perfect if the perturbation is absent ($\delta = 0$). For non-zero perturbations, OLE provides valuable information about the nature of operator spreading.

1. *OLE estimation protocol*

Suppose O is a Pauli string comprising only I- and Z-operators on a few qubits $\mathcal{Q} \subset \{0, \ldots, N-1\}$, $|\mathcal{Q}| \ll N$, i.e., $O = Z_\mathcal{Q} \equiv (\otimes_{q\in\mathcal{Q}} Z_q) \otimes (\otimes_{q\notin\mathcal{Q}} I)$. Then the experiment design is as follows:

1. prepare a randomly chosen bitstring-state $|\phi_s\rangle \in \{|0\ldots0\rangle, \ldots, |1\ldots1\rangle\}$;
2. calculate the state parity $\langle\phi_s| Z_\mathcal{Q} |\phi_s\rangle \in \{+1, -1\}$ on the subset $\mathcal{Q}$ of qubits;
3. evolve the state $|\phi_s\rangle$ in accordance with the unitary evolution $U_\delta(t) U_0^\dagger(t)$;
4. measure the evolved state $|\phi_s(t)\rangle$ in the computational basis so as to estimate $\langle\phi_s(t)| Z_\mathcal{Q} |\phi_s(t)\rangle$;
5. repeat the steps 1 to 4 S times;
6. calculate the estimator L_S for the operator Loschmidt echo and its estimation error ΔL_S as the average of products $\langle\phi_s| Z_\mathcal{Q} |\phi_s\rangle \times \langle\phi_s(t)| Z_\mathcal{Q} |\phi_s(t)\rangle$ and its standard error.

In strike contrast to the Loschmidt echo for states, the operator Loschmidt echo does not require any measurements to be performed on the qubits beyond the subset $\mathcal{Q}$. This results in the exponentially higher frequency of events f_{OLE} when the values $\langle\phi_s| Z_\mathcal{Q} |\phi_s\rangle$ and $\langle\phi_s(t)| Z_\mathcal{Q} |\phi_s(t)\rangle$ correlate, namely, $f_{\mathrm{OLE}} \sim e^{-\delta|\mathcal{Q}|t} \gg e^{-\delta N t} \sim f_{\mathrm{LES}}$.

2. *Multiple-OLE estimation protocol*

Should all N qubits be measured in the standard basis as an output of quantum computation anyway, this opens an option to specify the target subset of qubits $\mathcal{Q}$, and consequently the observable $Z_\mathcal{Q}$, *a posteriori*. This specifics of quantum computation unlocks its potential of estimating multiple OLEs at once, which is prohibitive to many classical simulation methods. The experiment design for addressing multiple OLEs is as follows:

1. prepare a randomly chosen bitstring-state $|\phi_s\rangle \in \{|0\ldots0\rangle, \ldots, |1\ldots1\rangle\}$;

2. evolve the state $|\phi_s\rangle$ in accordance with the unitary evolution $U_\delta(t)U_0^\dagger(t)$;
3. measure the evolved state $|\phi_s(t)\rangle$ in the computational basis and save the measurement-outcome bitstring m_s along with the initial state $|\phi_s\rangle$;
4. repeat the steps 1 to 3 S times;
5. choose a target subset of qubits $\mathcal{Q}$ for OLE;
6. calculate the parity $\langle\phi_s| Z_\mathcal{Q} |\phi_s\rangle \in \{+1, -1\}$ for each s;
7. calculate $\langle m_s| Z_\mathcal{Q} |m_s\rangle$ for each s;
8. calculate the estimator $L_S(Z_\mathcal{Q})$ for the target operator Loschmidt echo and its estimation error $\Delta L_S(Z_\mathcal{Q})$ as the average of products $\langle\phi_s| Z_\mathcal{Q} |\phi_s\rangle \times \langle m_s| Z_\mathcal{Q} |m_s\rangle$ and its standard error;
9. repeat the steps 5 to 8 for all target observables $Z_\mathcal{Q}$.

D. OLE and OTOC

In case the δ-perturbation is time-localised in the middle of echo dynamics, i.e., $U_\delta(t) = U_0(t)V_\delta$, then the OLE is directly related with the out-of-time-order correlation function (OTOC) [22, 24, 25]. The OLE in Eq. (21) takes the form

$$L = \frac{\mathrm{tr}\big(O_0^\dagger(t)V_\delta O_0(t)V_\delta^\dagger\big)}{\mathrm{tr}\,(O^\dagger O)}. \tag{27}$$

Let $G = G^\dagger$ be a generator of the perturbation V_δ, i.e., $V_\delta = e^{-i\delta G}$. Then we have

$$\begin{aligned} V_\delta O_0(t) V_\delta^\dagger &= e^{-i\delta G} O_0(t) e^{i\delta G} \\ &= O_0(t) - i\delta[G, O_0(t)] - \frac{\delta^2}{2}[G,[G,O_0(t)]] \\ &\quad + \ldots + \frac{(-i\delta)^k}{k!}\underbrace{[G,[G,[\ldots[G,O_0(t)]\ldots]]]}_{k\text{th-order commutator}} + \ldots \end{aligned} \tag{28}$$

Substituting (28) in (27) and taking into account the relation $\mathrm{tr}(A[B,C]) = \mathrm{tr}([A,B]C)$, we obtain

$$L = 1 + i\delta \times \frac{\mathrm{tr}\big([O_0^\dagger(t), O_0(t)]G\big)}{\mathrm{tr}\,(O^\dagger O)} - \frac{\delta^2}{2} \times \frac{\mathrm{tr}\big([G,O_0(t)]^\dagger[G,O_0(t)]\big)}{\mathrm{tr}\,(O^\dagger O)} + \ldots \tag{29}$$

For an Hermitian operator $O = O^\dagger$, the first-order term $\propto \delta$ vanishes. So do all odd-order terms because

$$\mathrm{tr}\Big(O_0^\dagger(t)\underbrace{[G,[G,[\ldots[G,O_0(t)]\ldots]]]}_{k\text{th-order commutator}}\Big) = (-1)^k \mathrm{tr}\Big(O_0(t)\underbrace{\big[G,[G,[\ldots[G,O_0^\dagger(t)]\ldots]]\big]}_{k\text{th-order commutator}}\Big).$$

Therefore, the OLE for an Hermitian traceless operator O with eigenvalues ± 1 reads

$$\begin{aligned} L &= 1 - \frac{\delta^2}{2} \times \underbrace{\frac{1}{2^N}\mathrm{tr}\big([G,O_0(t)]^\dagger[G,O_0(t)]\big)}_{\mathrm{OTOC}} \\ &\quad + \frac{\delta^4}{4!} \times \underbrace{\frac{1}{2^N}\mathrm{tr}\big([G,[G,O_0(t)]]^\dagger[G,[G,O_0(t)]]\big)}_{\geqslant 0} + \ldots \end{aligned} \tag{30}$$

The terms in the second line of Eq. (30) are negligible provided $\delta \ll 1$. The validity of omitting those terms is probed experimentally: the assumption is justified as long as the dependence of OLE on δ^2 remains linear in the vicinity of $\delta = 0$ and does not bend up due to the positive fourth-order correction. The slope in the linear dependence

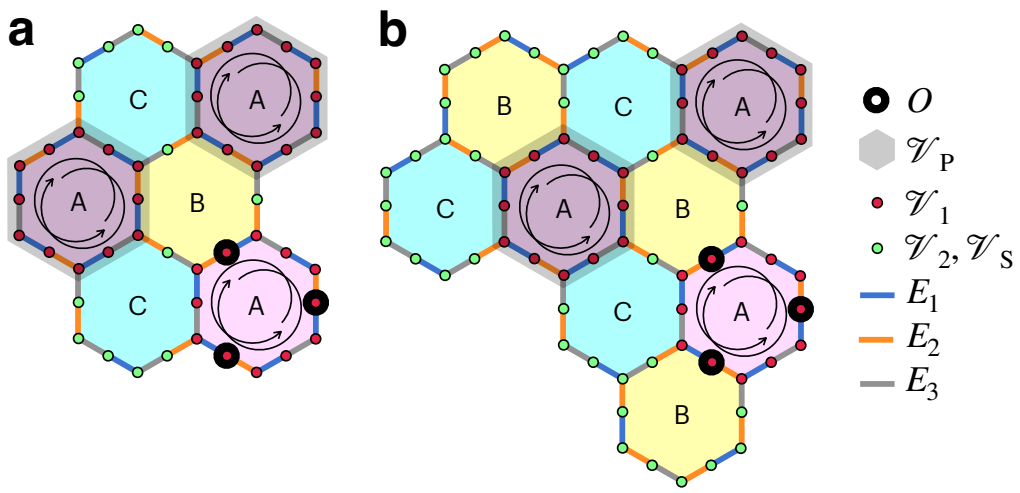

FIG. 4. **Heterogeneous structures probed in the experiment.**

$\mathrm{OLE}(\delta^2) = 1 - \frac{1}{2}\delta^2 \times \mathrm{OTOC}$ enables one to reveal the OTOC in the maximally mixed (infinite temperature) state $\varrho_\infty = (\frac{1}{2}I)^{\otimes N}$,

$$\mathrm{OTOC} = \langle [G, O_0(t)]^\dagger [G, O_0(t)] \rangle_{\varrho_\infty}. \tag{31}$$

The OTOC can also be seen as an ensemble average over the Loschmidt echos for states [26, 27]. However, the OLE provides a more practical way to estimate the OTOC in the presence of noise.

V. EXPERIMENTAL DESIGN AND THE CASE CIRCUITS

The experimental design designed by Algorithmiq is comprehensively outlined in Fig. 2 as a summary of the previous sections. Global rescaling with respect to $\delta = 0$ is used as the simplest noise-agnostic mitigation method. The noise-aware methods such as the tensor-network error mitigation (TEM) [28–30] relying on the learnt structure of the noise are being tested on complex 2D topologies.

Insofar as the experiment concerned, the first demonstration is done on the heavy-hex topology for the structures depicted in Fig. 4 with all the key vertices and edges indicated. The model given by Eqs. (8) and (9) is used with the parameters $h = \frac{\pi}{8}$, $b_1 = \frac{3\pi}{16}$, $b = 0.25$ for $L = 3, 6$. The experiment is run by IBM's team.

VI. CLASSICAL SIMULATIONS

A. Belief-propagation tensor-network simulations

- Belief propagation is first and foremost a contracting self-consistent algorithm for tensor network. Its efficacy is essential for looped tensor networks, especially for large loops in which the exact tensor contraction often assumes prohibitive costs, and it is where belief propagation actually work best.
- Being a contracting method, its original definition is on closed tensor Network with no open *legs*. In this framework, the algorithm can be summarized as follows:
  1. For each edge in the network, define two starting guess messages (vectors) of dimension as the bond dimension of the edge, one for each direction of the edge $i \to j$, $j \to i$, with i, j being the tensors surrounding the edge in consideration.
  2. For each iteration of the algorithm, each message $i \to j$ is defined as the contraction of the tensor i with all the messages directed to i ($x \to i$). This is the update rule of the self-consistent algorithm, it can be applied in parallel over each message at the same time, or sequentially, one message at a time.
  3. Apply the update of the messages until some convergence criterion is met.

4. Now, each tensor contraction with its incoming messages represents a valid contraction value of the network, but to account for differences in convergence across the network, one usually applies some kind of averaging over all the possible contraction values. This is often expressed in terms of a partition function (see [31]) in analogy with physics.

- The quality of the self consistent BP approximation can be checked by means of the loop correlation analysis, where each loop (or only some of them) is broken at a chosen edge, and by inserting messages over the indices reaching the loop from the rest of the network, one can analyze the eigenvalues of the matrix resulting of the now linearized and contracted loop [17]. This can give a clear idea of how well the messages approximate the contraction value.

- For tougher cases of numerous tight loops, such as on a square lattice topology of the network, one can resort to the loop series expansion [31], which uses the notion that loop correlation decays exponentially with the loop size [32] to average over bigger loops of the network. This is done by inserting the messages over a partition of the network containing more than one tensor, and resorting to exact contraction within the same partition.

- While, as we explained, belief-propagation is in first instance a contracting method, it has been recently shown to reliably provide the Vidal-gauge of a tensor-network-state in which belief propagation is applied to the network formed by the pair of the sites $i, i^\dagger$ joint together [33]. This property makes it a remarkable tool not only to contract a network, but also to evolve it via application of *gates*, a.k.a. to simulate the evolution of the network.

B. Pauli propagation method

The OTOC quantity in Eq. (31) can be efficiently computed using the Pauli propagation framework [34–36]. Given an initial Pauli string O_0 and a gate-based unitary evolution U, we apply the constituent gates sequentially while tracking the generated Pauli strings and their corresponding coefficients. The gates comprising the unitary evolution are Pauli rotations of the form $R_Q(\theta) = \exp(-i\theta Q/2)$, where Q is the Pauli string that generates the rotation.

When applying the gate $R_Q(\theta)$ sequentially to all Pauli strings in the decomposition of operator $O(t)$ from Eq. (18), new Pauli strings and coefficients are generated according to the transformation rule:

$$R_Q^\dagger(\theta) P R_Q(\theta) = \begin{cases} P, & \text{if } [P,Q] = 0 \\ \cos(\theta)P + i\sin(\theta)P', & \text{if } \{P,Q\} = 0, \end{cases} \tag{32}$$

where $P' = i[P,Q]/2$. Thus, when the generator Q of the rotation does not commute with the existing Pauli string P, a new Pauli string is generated through the anticommutation relation. For highly scrambling circuits with substantial non-Clifford character, the exponential growth in the number of Pauli strings renders exact tracking computationally intractable.

To address this, we implement truncation strategies that discard Pauli strings. Two primary truncation rules are employed: *coefficient truncation*, whereby Pauli strings with coefficients below a predefined threshold are discarded, and *weight truncation*, which eliminates Pauli strings when $w(P)$ exceeds a specified maximum value [8].

In practice, using 32 GB of available memory with these truncation methods, we encountered two possible outcomes: either the signal decays to zero due to excessive Pauli string elimination, or the simulation exhausts available memory resources. This limitation motivates the exploration of alternative simulation and truncation approaches.

To quantify the computational resources required for accurate Pauli propagation simulations, we employ the following estimation procedure. Using the single-path Pauli simulations described in Sec. VI C, we first compute the typical weight w_m of Pauli strings comprising the observable $O(t)$ in its Pauli decomposition. The value w_m has been empirically observed to depend on the parameters of the circuit: number of Floquet layers L, total number of qubits N_q, and the parameter b. By identifying the median weight value w_m given specific experimental parameters, we can estimate the memory requirement in gigabytes for faithful representation of the evolution as

$$M = \binom{N_q}{w_m} 3^{w_m} \frac{(b_p + b_c)}{8 \times 10^9}, \tag{33}$$

where N_q is the number of qubits, b_p and b_c are the bits required to represent one Pauli string and its coefficient, respectively, and 8×10^9 provides the conversion factor from bits to gigabytes. The binomial coefficient $\binom{N_q}{w_m}$ accounts for the number of ways to distribute the non-identity Pauli operators among N_q qubits, while 3^{w_m} represents the choice of Pauli matrices (X, Y, Z) at each non-identity position.

Representative examples of memory requirements are presented in Table I, demonstrating the exponential scaling that renders exact Pauli propagation intractable for moderate system sizes and circuit depths.

N_q	L	b	$b_p + b_c$	w_m	Memory (GB)
49	3	0.25	192	19	10^{14}
49	6	0.25	192	34	10^{20}
70	3	0.25	320	21	10^{19}
70	6	0.25	320	39	10^{31}

TABLE I. Memory requirements for simulating quantum evolution using exact Pauli propagation. The parameters L denotes the number of Floquet layers, b represents the perturbation strength, and w_m is the median Pauli weight. The calculation assumes that all Pauli strings must be retained during evolution, as it is not possible to determine *a priori* which strings will contribute significantly to the final observable.

C. Monte Carlo Methods

An alternative computational strategy involves Monte Carlo sampling of Pauli strings to reconstruct $O(t)$, and use it to compute the commutator with generator G in Eq. (31). Let N_P denote the maximum number of Pauli strings retained in memory during the simulation. Beginning with the single initial Pauli string O_0, the application of non-Clifford gates generates an expanding collection of Pauli strings according to the evolution rule in Eq. (32).

When the number of tracked Pauli strings approaches the imposed limit N_P, exact enumeration of all newly generated strings becomes impossible. Then, when a gate application would generate two Pauli strings following Eq. (32), we employ probabilistic sampling to retain only one string. The selection follows the natural probability distribution determined by $\cos^2(\theta)$ and $\sin^2(\theta)$, respectively, corresponding to the squared amplitudes of the transformation coefficients.

The normalization of coefficients in the evolving Pauli sum is maintained throughout the evolution process, though only the resulting Pauli strings are collected at the end of each run. Upon completion of S Monte Carlo realizations, we obtain a collection $\{n_P\}$ representing the number of appearances for each distinct Pauli string across all runs.

To estimate the evolved operator $O(t)$, we compute the coefficients in the Pauli decomposition as $c_P \propto 1/\sqrt{n_P}$, ensuring the normalization condition $\|O(t)\|_2 = 1$. The commutator $[G, O(t)]$ can then be directly evaluated, yielding the OTOC value specified in Eq. (31).

Empirically, optimal performance is achieved when restricting the evolution to a single Pauli string ($N_P = 1$). However, as the number of simultaneously tracked Pauli strings increases beyond unity, the estimated OTOC value systematically decreases toward zero, indicating the accumulation of sampling errors in the Monte Carlo approximation.

While this method is computationally efficient in terms of resource requirements, it suffers from a fundamental limitation. The simulation protocol is insensitive to the sign of newly generated Pauli strings, as the sampling decision is based solely on the squared amplitudes $\cos^2(\theta)$ and $\sin^2(\theta)$. Consequently, the substitution $\theta \to -\theta$ in any Pauli rotation leaves the Monte Carlo sampling outcome unchanged, rendering the method unable to distinguish between unitaries that differ only in the signs of their rotation angles. This phase insensitivity represents a significant constraint on the method's applicability to problems where interference effects are crucial.

D. Statistical approach: Full-scrambling assumption

Assuming the dynamics is fully scrambling, observable $O(t)$ is composed of random Pauli strings, and the distribution of the squares of the strings' coefficients is uniform. Then

$$\mathrm{OTOC} = 2|\mathscr{V}_P| \tag{34}$$

because every Pauli component commutes or anticommutes with the perturbation generator G with probability $\frac{1}{2}$. Each commutator in formula (31) results in an additional multiplier 2.

[1] J. Eisert and J. Preskill, Mind the gaps: The fraught road to quantum advantage (2025), arXiv:2510.19928 [quant-ph].
[2] Z. Zimborás, B. Koczor, Z. Holmes, E.-M. Borrelli, A. Gilyén, H.-Y. Huang, Z. Cai, A. Acín, L. Aolita, L. Banchi, F. G. S. L. Brandão, D. Cavalcanti, T. Cubitt, S. N. Filippov, G. García-Pérez, J. Goold, O. Kálmán, E. Kyoseva, M. A. C. Rossi, B. Sokolov, I. Tavernelli, and S. Maniscalco, Myths around quantum computation before full fault tolerance: What no-go theorems rule out and what they don't (2025), arXiv:2501.05694 [quant-ph].

[3] J. Tindall, M. Fishman, E. M. Stoudenmire, and D. Sels, Efficient tensor network simulation of IBM's eagle kicked ising experiment, PRX Quantum **5**, 010308 (2024).
[4] T. Begušić and G. K.-L. Chan, Fast classical simulation of evidence for the utility of quantum computing before fault tolerance (2023), arXiv:2306.16372 [quant-ph].
[5] T. Begušić, J. Gray, and G. K.-L. Chan, Fast and converged classical simulations of evidence for the utility of quantum computing before fault tolerance, Science Advances **10**, eadk4321 (2024).
[6] P. Rall, D. Liang, J. Cook, and W. Kretschmer, Simulation of qubit quantum circuits via pauli propagation, Phys. Rev. A **99**, 062337 (2019).
[7] P. Bermejo, P. Braccia, M. S. Rudolph, Z. Holmes, L. Cincio, and M. Cerezo, Quantum convolutional neural networks are (effectively) classically simulable (2024), arXiv:2408.12739 [quant-ph].
[8] M. S. Rudolph, T. Jones, Y. Teng, A. Angrisani, and Z. Holmes, Pauli propagation: A computational framework for simulating quantum systems (2025), arXiv:2505.21606 [quant-ph].
[9] A. Miller, Z. Holmes, Özlem Salehi, R. Chakraborty, A. Nykänen, Z. Zimborás, A. Glos, and G. García-Pérez, Simulation of fermionic circuits using majorana propagation (2025), arXiv:2503.18939 [quant-ph].
[10] M. A. Nielsen and I. L. Chuang, *Quantum computation and quantum information* (Cambridge university press, 2010).
[11] 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.*, Quantum supremacy using a programmable superconducting processor, Nature **574**, 505 (2019).
[12] Observation of constructive interference at the edge of quantum ergodicity, Nature **646**, 825 (2025).
[13] Y. Kim, A. Eddins, S. Anand, K. X. Wei, E. van den Berg, S. Rosenblatt, H. Nayfeh, Y. Wu, M. Zaletel, K. Temme, and A. Kandala, Evidence for the utility of quantum computing before fault tolerance, Nature **618**, 500 (2023).
[14] X. Mi, P. Roushan, C. Quintana, S. Mandra, J. Marshall, C. Neill, F. Arute, K. Arya, J. Atalaya, R. Babbush, *et al.*, Information scrambling in quantum circuits, Science **374**, 1479 (2021).
[15] N. I. Zheludev and Y. S. Kivshar, From metamaterials to metadevices, Nature materials **11**, 917 (2012).
[16] S. Trebst and C. Hickey, Kitaev materials, Physics Reports **950**, 1 (2022), kitaev materials.
[17] M. S. Rudolph and J. Tindall, Simulating and sampling from quantum circuits with 2d tensor networks (2025), arXiv:2507.11424 [quant-ph].
[18] OLE experiment (2025).
[19] T. Zhou and B. Swingle, Operator growth from global out-of-time-order correlators, Nature communications **14**, 3411 (2023).
[20] T. Gorin, T. Prosen, T. H. Seligman, and M. Žnidarič, Dynamics of loschmidt echoes and fidelity decay, Physics Reports **435**, 33 (2006).
[21] T. Prosen, T. H. Seligman, and M. Žnidarič, Theory of quantum loschmidt echoes, Progress of Theoretical Physics Supplement **150**, 200 (2003), https://academic.oup.com/ptps/article-pdf/doi/10.1143/PTPS.150.200/5243181/150-200.pdf.
[22] M. Campisi and J. Goold, Thermodynamics of quantum information scrambling, Phys. Rev. E **95**, 062127 (2017).
[23] S. Karch, S. Bandyopadhyay, Z.-H. Sun, A. Impertro, S. Huh, I. P. Rodríguez, J. F. Wienand, W. Ketterle, M. Heyl, A. Polkovnikov, I. Bloch, and M. Aidelsburger, Probing quantum many-body dynamics using subsystem Loschmidt echos (2025), arXiv:2501.16995 [cond-mat.quant-gas].
[24] A. I. Larkin and Y. N. Ovchinnikov, Quasiclassical method in the theory of superconductivity, Journal of Experimental and Theoretical Physics (1969).
[25] J. Maldacena, S. H. Shenker, and D. Stanford, A bound on chaos, Journal of High Energy Physics **2016**, 106 (2016).
[26] B. Yan, L. Cincio, and W. H. Zurek, Information scrambling and loschmidt echo, Phys. Rev. Lett. **124**, 160603 (2020).
[27] Y.-N. Zhou and C. Liu, Generalized loschmidt echo and information scrambling in open systems (2024), arXiv:2412.01851 [quant-ph].
[28] S. Filippov, M. Leahy, M. A. C. Rossi, and G. García-Pérez, Scalable tensor-network error mitigation for near-term quantum computing (2023), arXiv:2307.11740 [quant-ph].
[29] S. N. Filippov, S. Maniscalco, and G. García-Pérez, Scalability of quantum error mitigation techniques: from utility to advantage (2024), arXiv:2403.13542 [quant-ph].
[30] L. E. Fischer, M. Leahy, A. Eddins, N. Keenan, D. Ferracin, M. A. C. Rossi, Y. Kim, A. He, F. Pietracaprina, B. Sokolov, S. Dooley, Z. Zimborás, F. Tacchino, S. Maniscalco, J. Goold, G. García-Pérez, I. Tavernelli, A. Kandala, and S. N. Filippov, Dynamical simulations of many-body quantum chaos on a quantum computer (2024), arXiv:2411.00765 [quant-ph].
[31] J. Gray, G. Park, G. Evenbly, N. Pancotti, E. F. Kjønstad, and G. K.-L. Chan, Tensor network loop cluster expansions for quantum many-body problems (2025), arXiv:2510.05647 [quant-ph].
[32] S. Midha and Y. F. Zhang, Beyond belief propagation: Cluster-corrected tensor network contraction with exponential convergence (2025), arXiv:2510.02290 [quant-ph].
[33] J. Tindall and M. Fishman, Gauging tensor networks with belief propagation, SciPost Phys. **15**, 222 (2023).
[34] A. Angrisani, A. Schmidhuber, M. S. Rudolph, M. Cerezo, Z. Holmes, and H.-Y. Huang, Classically estimating observables of noiseless quantum circuits (2025), arXiv:2409.01706 [quant-ph].
[35] A. Angrisani, A. A. Mele, M. S. Rudolph, M. Cerezo, and Z. Holmes, Simulating quantum circuits with arbitrary local noise using pauli propagation (2025), arXiv:2501.13101 [quant-ph].
[36] M. S. Rudolph, E. Fontana, Z. Holmes, and L. Cincio, Classical surrogate simulation of quantum systems with lowesa (2023), arXiv:2308.09109 [quant-ph].