Temporal entropy and the complexity of computing the expectation value of local operators after a quench

Stefano Carignano [1,*] Carlos Ramos Marimón [2] and Luca Tagliacozzo [3,†]

[1]*Barcelona Supercomputing Center, 08034 Barcelona, Spain*

[2]*Departament de Física Quàntica i Astrofísica and Institut de Ciències del Cosmos (ICCUB), Universitat de Barcelona, Martí i Franquès 1, 08028 Barcelona, Catalonia, Spain*

[3]*Institute of Fundamental Physics IFF-CSIC, Calle Serrano 113b, Madrid 28006, Spain*

We study the computational complexity of simulating the time-dependent expectation value of a local operator in a one-dimensional quantum system by using temporal matrix product states(MPSs) and argue that it is intimately related to that of encoding temporal transition matrices and their partial traces. We show that we can upper-bound the rank of these reduced transition matrices by one of the Heisenberg evolution of local operators, thus making aconnection between two apparently different quantities, the temporal entanglement and the local operator entanglement(OE). As a result, whenever the local OE grows slower than linearly in time, we show that computing time-dependent expectation values of local operators using temporal MPSs is likely advantageous with respect to computing the same quantities using standard MPS techniques.

I. INTRODUCTION

The complexity of simulating quantummany-body systems increases exponentially with the number of constituents. Over the last few decades, the development of tensor network (TN) techniques has, however, helped in gaining better insight into the equilibrium properties of many-body quantum systems. It is now understood that, at equilibrium, quantum complexity is mostly related to the amount of entanglement, and we have been able to design TN *Ansätze* that can encode the structure of typical equilibrium states.

Out of equilibrium, the situation is different: Even in the simplest protocol, e.g., quantum quenches, correlations spread over large distances, quickly producing robustly entangled states [1]. Standard TN techniques such as those based on matrix product states (MPSs) thus struggle to cope with the fast growth of entanglement, and as a result, their cost increases exponentially with the duration of the evolution [2]: This is often referred to as the *entanglement barrier* in the literature [3].

This fact is a consequence of trying to represent the full quantum state during the evolution. The situation might change if we focus on a local description of the state, by trying to describe the evolution of the expectation value of local operators. In principle, this is a much simpler task, and several approaches along these lines have been proposed

*stefano.carignano@bsc.es

†luca.tagliacozzo@iff.csic.es

[4–12]. However, except for few studies on integrable systems [13,14], there has been no systematic understanding of the real computational cost of such approaches and thus no concrete understanding of the complexity of simulating the evolution of local observables.

In this paper, we make a step in this direction. We consider the evolution under a local Hamiltonian of the expectation value of a local observable and follow the inspirational papersin Refs. [4–6,10,12]. These algorithms stillrely on a MPS, which is however now defined in time, making it a temporal MPS (tMPS) along a Keldysh contour [6,15]. It has recently been noticed that such tMPSs encode the influence matrix of the system [16] that drives the evolution of a region of the full system and thus provides the systematic way of translating the global evolution into a local one [11,17,18]. In specific cases, such tMPSs can be described with small bond dimension, though this is not always the case [4,5].

Here, we provide some theoretical backup to these numerical observations. Our main result in this direction is to bound the complexity of the tMPS with the one of encoding the time evolution of operators in the Heisenberg picture. Such complexity is encoded in the so-called *operator entanglement* (OE) [19,20], which is expected to grow at most logarithmically in time for integrable systems, whereas it increases linearly for nonintegrable systems [3,21–23]. The key to obtaining this result is a slightly modified version of the algorithms introduced in Refs. [4,6] which, as shown in the following, allowsus to build a direct connection between the two quantities.

Whenever the OE grows logarithmically in time, we can thus show that the tMPS bond dimension is bounded by a polynomial growth in time. On the other hand, when the OE grows linearly in time, our bound on the tMPS bond dimension is exponential. While this does not rule out that there
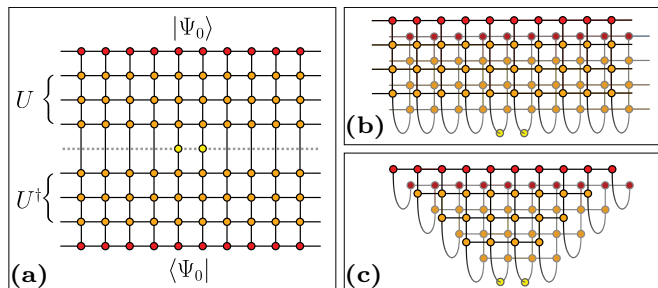
FIG. 1. (a) The time-dependent expectation value of a local operator acting on a one-dimensional system in the Keldysh representation can be described by (b) a double sheet two-dimensional (2D) tensor network (TN) contraction. (c) Upon using a Trotter approximation and the locality of the operator, it simplifies to a triangular TN.

might be specific cases in which the tMPS might still have a small bond dimension and thus provide a computational advantage, see, e.g., Refs. [22–24], we believe that these are fortuitous exceptions to the general rule. Indeed, in the scenarios we presented for this case as well as others in the literature [10], we find that sucha bound is saturated.

II. SETUP

Given a lattice system, we consider the complexity of computing the evolution of the expectation value of a local operator after a quench. We start from a product state $|\psi_0\rangle$ and a local Hamiltonian H. The system starts to evolve, and it is described by the state $|\psi(T)\rangle = \exp(-iHT)|\psi_0\rangle \equiv U(T)|\psi_0\rangle$. In general, the entanglement entropy of the state increases linearly in time [1], and standard time-dependent MPS simulations become exponentially expensive [25,26] (for recent reviews, see also Refs. [27,28]). Here, rather than evolving the state, we focus on the evolution of the expectation value of a local operator acting on two neighboring sites [29]:

$$\langle \mathcal{O}\mathcal{Q}\rangle (T) = \langle \psi(T)| \mathcal{O}_i \mathcal{Q}_{i+1} |\psi(T)\rangle. \quad (1)$$

Approximating $U(T)$ by a sequence of short evolutions $U(\delta t)^{N_T}$ with $N_T = T/\delta t$ and using a Trotter expansion, this quantity is encoded in the contraction of a two-dimensional TN containing order $N_x \times N_T$ tensors [30,31], see Fig. 1(a). Furthermore, one can fold the network following Refs. [4,5], leading to a double-layer structure, as shown in Fig. 1(b). Given the Trotter approximation, the time evolution of a local operator has an exact causal cone obtained by canceling all the unitary gates that are contracted with their respective Hermitian conjugates. As a result, the network can be simplified and acquires the triangular shape shown in Fig. 1(c) [10,12].

A. Compressing the tMPS

Being two-dimensional, the best contraction path for this TN is not *a priori* obvious. A possible strategy is to identify two tMPSs defining the contraction of the left and right halves of the system (see Fig. 2) and interpret the triangular network as the scalar product between the two $\langle \mathcal{O}\mathcal{Q}\rangle (t) = \langle L_\mathcal{O}|R_\mathcal{Q}\rangle$ [32]. The identification is purely formal since the bond dimension of the individual tMPS tensors can grow exponentially
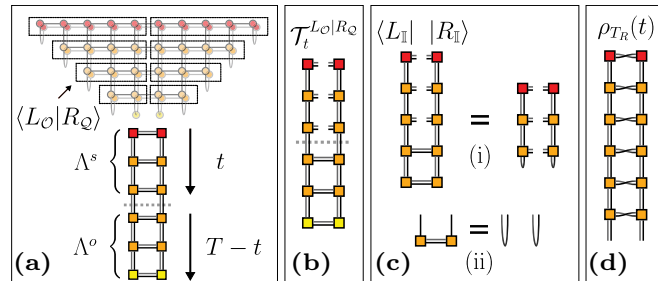

FIG. 2. (a) The triangular tensor network (TN) can be contracted from the sides, identifying left $\langle L_\mathcal{O}|$ and right $|R_\mathcal{Q}\rangle$ temporal matrix product states (tMPSs). The upper-most tensors of the tMPSs are dictated by the initial state, the lower-most ones by the choice of the operator. (b) The reduced transition matrix. (c) In the absence of operators, the folded tensors of the time evolution resolve to identities. (d) The partial transpose of the time-evolved left-right density matrix of the system, where forward and backward legs are swapped.

with the number of time steps. The construction of the tMPS thus only makes sense if one can show that, at least for specific scenarios, its bond dimension increases mildly with the number of time steps.

The task is therefore to identify the relevant rank of the tMPS matrices and compress them on their support. We know that the tMPS allowsus to compute correlation functions of the type $\langle \mathcal{O}\mathcal{P}(t) \cdots \mathcal{Q}\mathcal{R}(t')\rangle$ for an arbitrary number of insertions of local operators at different times between 0 and T. Following the standard density matrix renormalization group (DMRG) recipe [26], this requires having a faithful representation of all reduced transition matrices (RTM) $\mathcal{T}_t^{L_\mathcal{O}|R_\mathcal{Q}}$, defined as

$$\mathcal{T}_t^{L_\mathcal{O}|R_\mathcal{Q}} = \mathrm{Tr}_{T-t}[\mathcal{T}^{L_\mathcal{O}|R_\mathcal{Q}}], \quad \mathcal{T}^{L_\mathcal{O}|R_\mathcal{Q}} = \frac{|R_\mathcal{Q}\rangle \langle L_\mathcal{O}|}{\langle L_\mathcal{O}|R_\mathcal{Q}\rangle}, \quad (2)$$

which can be seen as generalizations of reduced density matrices.

The graphical representation of these objects is found in Fig. 2(b). Interestingly, such transition matrices have also been considered in the context of holography: The properties of the equivalent in the field theory of the transition matrices described here have a geometrical interpretation in the bulk [33–36]. It is then the rank of such RTMs which dictates the rank of the tMPS matrices rather than the temporal entropy that was previously considered [4,6,10,12].

Generalizing the DMRG prescription, the bond dimension of the tMPS tensors should allowus to obtain a low-rank approximation of the RTM $\mathcal{T}_t^{L_\mathcal{O}|R_\mathcal{Q}}$, $\forall t \in 0 \cdots T$, with the desired precision.

The RTMs are complex-valued, and for reflection-invariant Hamiltonians, whenever $\mathcal{Q} = \mathcal{O}$, they are symmetric. As a result, their low rank approximation is better defined in terms of their singular values, so we will project the tMPS bond dimension at time t on the largest singular values of $\mathcal{T}_t^{L_\mathcal{O}|R_\mathcal{Q}}$.

Following Ref. [6], we now define $\Lambda^s(t)$ and $\Lambda^o(t)$, respectively, as the contractions of the overlap $\langle L_\mathcal{O}|R_\mathcal{Q}\rangle$ until t (the top part of the network contraction, including the initial state) and below t (the bottom part, containing the operator), see

Fig. 2. We also define $\bar{\Lambda}_L^s(t)$, $\bar{\Lambda}_R^s(t)$ as the contraction up to t of the network obtained by contracting $\langle L_{\mathcal{O}} | L_{\mathcal{O}} \rangle$ and $\langle R_{\mathcal{Q}} | R_{\mathcal{Q}} \rangle$ [37].

With these definitions, we have that

$$\mathcal{T}_t^{L_{\mathcal{O}} | R_{\mathcal{Q}}} \simeq \sqrt{\bar{\Lambda}_L^s(t)} \Lambda^o(t) \sqrt{\bar{\Lambda}_R^s(t)}, \qquad (3)$$

where we use the similarity symbol to indicate that the two operators share the same singular values. For reflection-invariant systems, such as those we analyze here, one has $\bar{\Lambda}_R^s(t) = \bar{\Lambda}_L^s(t)$. Thus, our central result reads

$$\mathcal{R}\left(\mathcal{T}_t^{L_{\mathcal{O}} | R_{\mathcal{Q}}}\right) \leqslant \min\left\{\mathcal{R}\left[\bar{\Lambda}_L^s(t)\right], \mathcal{R}[\Lambda^o(t)]\right\}, \qquad (4)$$

where $\mathcal{R}$ denotes the rank of a given matrix. This equation is at the basis of our analysis.

Since for each choice of $\mathcal{O}$ and $\mathcal{Q}$ we obtain a different tMPS representation or $\langle L_{\mathcal{O}}(t)|$ and $|R_{\mathcal{Q}}(t)\rangle$, we define the cost of the algorithm as the cost of simulating the operators that require the highest bond dimension.

B. Rank of RTMs

Whenever $\mathcal{O} = \mathcal{Q} = \mathbb{I}$, the transition matrices $\mathcal{T}_t^{L_{\mathbb{I}} | R_{\mathbb{I}}}$ for all t are projectors. As a result, the states $|L_{\mathbb{I}}\rangle$ and $|R_{\mathbb{I}}\rangle$ consist of trivial singlets at the virtual level, as sketched in Fig. 2(c), implying that the tMPS matrices always have bond dimension 1, for all t. To obtain nontrivial tMPS, we thus include the operators in the construction of the states.

From the previous definitions, it follows that

$$\Lambda^o(t) = U(t)(\mathcal{O} \otimes \mathcal{Q}) U(t)^\dagger, \qquad (5)$$

meaning that $\Lambda^o(t)$ exactly encodes the Heisenberg evolution of the initially localized operator $\mathcal{O}_i \mathcal{Q}_{i+1}$, while

$$\Lambda^s(t) = \rho(t) = U(t) |\psi_0\rangle \langle\psi_0| U(t)^\dagger, \qquad (6)$$

where $\rho(t)$ is the time-evolved density matrix of the initial state. Finally, for Hamiltonians that are invariant under reflections with respect to the center of the chain,

$$\bar{\Lambda}_L^s(t) = \bar{\Lambda}_R^s(t) = \rho(t)_{T_R}, \qquad (7)$$

where T_R stands for the partial transpose on the semi-infinite right part of the system. Since we are evolving a pure state, there are known relations between $\rho_{T_R}(t)$ and $\rho(t)$ [38], and we know that $\mathcal{R}[\bar{\Lambda}_R^s(t)] = \mathcal{R}[\Lambda_R^s(t)]$ [39,40].

Having identified the elements composing the RTMs in Eq. (3), we can now use their physical properties together with Eq. (4) to identify some useful bound on the rank of the tMPS. A first scenario one can consider is that of a local quench. In this case, it is well known that the entropy of the evolved state only increases logarithmically with time, and as a result, the rank of $\Lambda^s(t)$ increases at most polynomially [41,42]. Thus, using Eq. (4), we have

$$\mathcal{R}\left(\mathcal{T}_t^{L_{\mathcal{O}} | R_{\mathcal{Q}}}\right) \leqslant T^\alpha, \qquad \forall t \in \{0, T\}. \qquad (8)$$

Such a case is perhaps not particularly interesting since a similar polynomial cost is obtained also by using standard MPS algorithms.

We thus turn to the more interesting scenario of a global quench. Here, we know that $\mathcal{R}[\Lambda^s(t)]$ increases exponentially

with T, as already mentioned. If now $\mathcal{R}[\Lambda^o(t)]$ only increases polynomially with T, we have

$$\mathcal{R}[\Lambda^o(t)] \leqslant t^\alpha \Rightarrow \mathcal{R}\left(\mathcal{T}_t^{L_{\mathcal{O}} | R_{\mathcal{Q}}}\right) \leqslant T^\alpha, \qquad \forall t, \qquad (9)$$

and the temporal MPS strategy provides a polynomial algorithm to compute the out-of-equilibrium dynamics of local observables, as already anticipated in Refs. [10,12]. In the literature, it is conjectured (and checked in several scenarios) that, in the case of integrable systems, the entanglement of local operators only grows logarithmically [3,19,20,22,23], implying that the rank of $\Lambda^o(t)$ only increases polynomially with T, as required in Eq. (9). As a result, if the conjecture is correct, the tMPS provides an efficient method to characterize the time evolution of local operators for integrable systems.

Since generically the rank of $\Lambda^o(t)$ is expected to grow exponentially with t, we expect that

$$\mathcal{R}\left(\mathcal{T}_t^{L_{\mathcal{O}} | R_{\mathcal{Q}}}\right) \leqslant \alpha \exp(\beta), \qquad (10)$$

with α and $\beta > 0$ model-dependent constants. In this case, there is no guarantee that the tMPS provides an efficient compression for the problem at hand.

C. Relation to other definitions of temporal entanglement

The basic ingredients in our formalism are the transition matrices defined in Eq. (2) and the generalized entanglement related to their rank. It is worth noting that these quantities differ from the definition of temporal entanglement recently proposed in the literature [4,6,10,12,13] which is built through density matrices obtained individually from either the left or right vectors (i.e., $\rho_R \sim |R\rangle \langle R|$, $\rho_L \sim |L\rangle \langle L|$).

This can be most easily seen in the case of a reflection-symmetric problem: Here, one has $|R\rangle = |\bar{L}\rangle$, so that the two definitions are connected by a nontrivial complex conjugation on the left state, which can be seen as a partial transpose operation acting on the legs of the folded transition matrix [see Fig. 2(d)]. This global unitary operation (the tensor product of swap gates on every physical leg, i.e., at each time site) acting on one side of the transition matrix can change the entanglement properties of the objects involved, so that a clear connection between the two quantities is not evident and would definitely require a deeper investigation in the future.

Another aspect which has been recently pointed out is that the traditional definition of temporal entanglement can depend on the gauge chosen to define the transverse transfer matrix and can be made arbitrarily small by using such gauge freedom. As a result, being able to accurately represent $|R\rangle$ and $\langle L|$ does not guarantee obtaining good precision on an expectation value $\langle O(T)\rangle = \langle L | O | R\rangle$ since the leading Schmidt vectors can be orthogonal to such an overlap [43]. Thus, as we argued, for our intent of determining the complexity of simulating expectation value of local operators, the use of the generalized entanglement proposed here is the appropriate object to consider, close in spirit to the concept of the biorthonormal truncation basis, which is already well established for non-Hermitian DMRG problems [43–45].

III. NUMERICAL RESULTS

We now report the numerical evidence to support our claims. We consider a transverse field Ising Hamiltonian:

$$H(g, h) = -\sum_i \left[\sigma_x^i \sigma_x^{i+1} + g\sigma_z^i + h\sigma_x^i \right], \qquad (11)$$

where $\sigma_{x,z}$ are Pauli matrices. We perform a second-order Trotter expansion of the time-evolution operator and cast it into a matrix product operator (MPO) following Ref. [31], with a timestep $\delta t = 0.05$. For our examples, we consider two cases: an integrable one for $g = 0.5$, $h = 0$, and a nonintegrable one $g = -1.05$, $h = 0.5$. We consider different initial product states, namely, $|0\rangle = (1, 0)^{\otimes N_x}$, $|+\rangle = (1, 1)/\sqrt{2}^{\otimes N_x}$, and $|r\rangle = (1, i)/\sqrt{2}^{\otimes N_x}$. The time evolution is performed by the variant of the standard folding algorithms we have described, exploiting the causal-cone of the network [4,6,10,12] and relying on a low-rank approximation of the RTMs defined in Eq. (2).

As in previous approaches [10,12], we iteratively construct the left and right vectors $\langle L_\mathcal{O}(T)|$ and $|R_\mathcal{Q}(T)\rangle$ starting from $\langle L_\mathcal{O}(T-1)|$ and $|R_\mathcal{Q}(T-1)\rangle$ by absorbing a new column of MPOs into them. These MPOs are obtained by contracting a column of the original TN, and they thus represent transfer matrices $\mathbb{E}^T$ evolving states of a time slice in space, a rotated version of the standard evolution in time. The bond dimension of the corresponding MPS increases, and we then keep it under control by projecting onto the largest D singular values of $\mathcal{T}_t^{L_\mathcal{O}|R_\mathcal{Q}}$ for every bipartition of the system into t and $T - t$. Further details about the algorithm and accurate comparison with other prescriptions in the literature are presented in the Appendixes.

In the case of the Ising model, we found the operator with the largest bond dimension defined at most on two sites to be a single-site σ_x. In the following, we will focus on optimizing with respect to it [46].

We start by considering the bond dimension necessary to keep the truncation error above a threshold $\epsilon = 10^{-4}$ in the singular value decomposition (SVD) spectrum of the RTMs and compare them with those obtained imposing the same truncation on the Heisenberg evolution for the operator $O(t)$, which entails the evolution of the vectorized local operator under the Hamiltonian $H \otimes \mathbb{I} - \mathbb{I} \otimes H$, resulting in the MPS $|\psi_O(t)\rangle$. The results are shown in Fig. 3, where we can observe that, as expected: (1) the behavior is different in the integrable and nonintegrable cases, and (2) the bond dimension necessary to correctly describe the RTMs are always below the ones necessary to describe the Heisenberg evolution of the operator.

We note, however, that to estimate the computational cost of simulating the expectation value of local operators with a finite precision using a tMPS, we should determine the bond dimension that is required to keep the distance in norm $|| \langle L_\mathcal{O}| - \langle L_\mathcal{O}^D| ||^2$ constant, where $\langle L_\mathcal{O}^D|$ is the truncation of the tMPS $\langle L_\mathcal{O}|$ to a given bond dimension D. Using the reasoning in Ref. [42], the term to consider would then be the overlap $\langle L_\mathcal{O}|L_\mathcal{O}^D\rangle$, which is upper bounded by the sum of the norm of the residuals discarded in the truncation process.

So far, we have just provided a bound on the overlap $\langle L_\mathcal{O}|R_\mathcal{O}\rangle$ which, even for reflection-invariant systems and
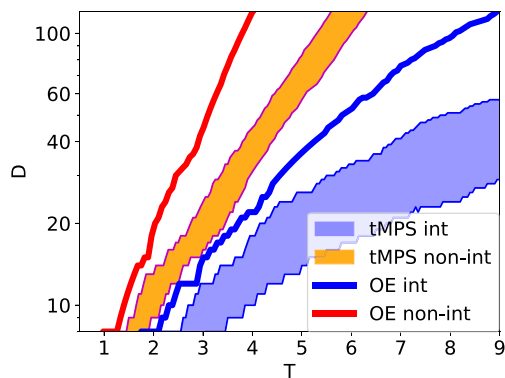

FIG. 3. Left: Bond dimensions D obtained by imposing a truncation error of 10^{-4} in the singular value decomposition of the reduced transition matrices (RTMs), as function of time (see main text). The shaded areas are delimited by the minimum and maximum D we found by varying the initial state. The thick solid lines represent the bond dimension curves for the operator entanglement (OE): They lie consistently above the corresponding ones for the temporal matrix product states (tMPSs). Blue denotes the integrable case (int), orange-red the nonintegrable one (non-int).

considering the symmetric case $\mathcal{O} = \mathcal{Q}$, differs from $\langle L_\mathcal{O}|L_\mathcal{O}^D\rangle$ by the absence of a complex conjugation. However, the two quantities are related since $\langle L_\mathcal{O}|L_\mathcal{O}^D\rangle$ can be obtained from $\langle L_\mathcal{O}|R_\mathcal{O}\rangle$ by evolving $|R_\mathcal{O}\rangle$ with one layer of local unitaries, representing the partial transpose (swapping the bra with the ket legs). Such an operation does not modify the rank of each state, but it reduces the overlap between the two, so that a shift in the resulting singular values can be expected.

As a result, the computational cost required for simulating the expectation value of a local operator with finite precision is constantly offset from that of computing $\langle L_\mathcal{O}|R_\mathcal{O}\rangle$, the latter being upper-bounded by the cost of the Heisenberg evolution of the operator. The scaling in time of the cost to maintain a constant error in local observables is the same as the one of $\langle L_\mathcal{O}|R_\mathcal{O}\rangle$ (see also the discussion in the Appendixes).

We verify this explicitly by following a procedure like the one proposed in Ref. [19] to estimate the truncation error in our tMPS compared with an exact result. The resulting bond dimensions for a fixed truncation error are reported in Fig. 4. For the integrable case, we find a polynomial increase, consistent with the behavior of the OE. Different initial states require different bond dimensions for a faithful description, though the growth follows the same power law. For the nonintegrable case, on the other hand, the OE requires an exponentially growing bond dimension for a faithful description, and the tMPS bond dimension follows the same behavior, consistent with our result.

IV. CONCLUSIONS

We have provided a connection between two separate concepts: temporal entanglement and the OE. We have shown that the rank of the RTM necessary to compute the temporal entanglement is upper-bounded by the rank of the OE of a bipartition. As a result, the scaling in time of the computational cost of simulating the evolution of a local observable with constant error is upper-bounded by the scaling of the OE, even
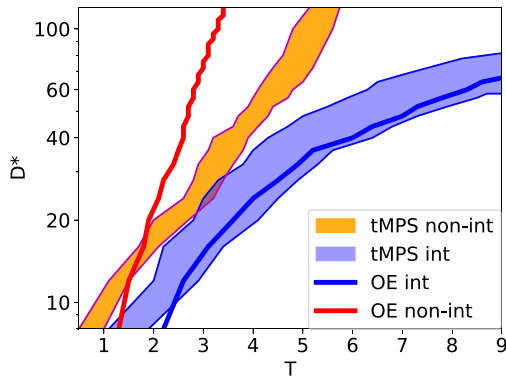
FIG. 4. Bond dimensions D^* obtained by imposing a 10^{-4} fidelity for the temporal matrix product state (tMPS) $\langle L_{\mathcal{O}}|$; again the shaded area is delimited by the minimum and maximum bond dimensions of the tMPS for a given initial state, and the thick solid lines denote the operator entanglement (OE; blue: integrable case, red: nonintegrable).

though the exact value of the cost can be offset by a constant value whose origin has been discussed in detail. Our presented algorithm yields the most efficient performance observed thus far, with a marginal advantage over previous approaches.

As a result, we can claim that, whenever the OE only grows logarithmically, using a tMPS for simulating the evolution of a local observable is the best choice in terms of computational cost and scales only polynomially with time, whereas in the generic case, the entanglement barrier cannot be circumvented by using the tMPS. On the other hand, a linear growth of the OE does not necessarily imply that the tMPS cannot provide an efficient compression, see, e.g., Refs. [22–24]. The transverse contraction methods discussed here would then be useful even in those cases.

Our results are a step toward understanding the cost of simulating the out-of-equilibrium dynamics of quantum many-body systems. It would be interesting to check the bounds we have obtained on a larger class of integrable and nonintegrable models and, more importantly, to explore if the transverse contraction can help in simulating the dynamics of higher dimensional systems.

Note added. While working on this paper, we found out about Ref. [48] discussing similar issues and suggesting a connection between the OE and the temporal entanglement, a connection we hope to have elucidated with our work.

APPENDIX A: THE ALGORITHM

In this appendix, we describe the algorithms we employ to obtain the left and right tMPSs encoding the contraction of the TN associated with the expectation value $\langle \mathcal{O}(T) \rangle$, where $\mathcal{O}$ is a local (or a set of local) operator acting on one or a few neighboring sites. In practice, for the iterative methods we employ, it is convenient to incorporate the operator(s) on one side of the network, as will become clear in the following. If we choose, e.g., to incorporate $\mathcal{O}$ on the right side of the network, our algorithms allow us to build $\langle L_{\mathbb{I}}|$ and $|R_{\mathcal{O}}\rangle$ for a given time T.

The basic building blocks of the TN are the MPO tensors representing the Trotterized time evolution (here, we employ the parameterization from Ref. [31]). For a given T which determines the extension of the network in the temporal direction ($N_T = T/\delta t$, δt being the Trotter step), we contract them in columns to generate the transfer matrices encoding the rotated (spacelike) evolution $\mathbb{E}^T$ for the folded system of which $\langle L|$ and $|R\rangle$ are the leading left and right eigenvectors in the thermodynamic limit. Depending on what we want to construct, we can also include the local operator at the bottom of the transfer matrix; in that case, we label it as $\mathbb{E}^T_{\mathcal{O}}$.

We have implemented both a power method for extracting the leading eigenvectors of $\mathbb{E}^T$ for a fixed time T as well as an algorithm based on the strict causal cone of the network when dealing with local operators. Both our algorithms are small variations of those presented in Refs. [4,10,12], the main difference being the cost function we use to perform the truncation required in the various iterations: We project the tMPS matrices on the support of the largest singular values of the respective RTM, as will be explained in the following.

The *power method* [see Fig. 5(a)] works by repeatedly applying the transfer matrix to an initial guess tMPS of length N_T until convergence is reached. More specifically, we start from the left $\langle L_{\mathbb{I}}|$ and right $|R_{\mathbb{I}}\rangle$ vectors and apply columns $\mathbb{E}^T$ to the left and $\mathbb{E}^T_{\mathcal{O}}$ to the right. At each step, the bond dimension of the tMPS increases by a factor d^2, where d is the physical dimension of the system constituents. Thus, to proceed, we truncate following our prescription and take the updated $\langle L_{\mathbb{I}}|$ as the input for the next step. With our parameterization of the MPO tensors, $\mathbb{E}^T$ is symmetric in left-right legs, so we can use $\langle L_{\mathbb{I}}|$ also as the new $|R_{\mathbb{I}}\rangle$. Alternatively, the optimization for $|R_{\mathbb{I}}\rangle$ simply requires an analogous step involving $\langle L_{\mathcal{O}}|$.

To determine whether the power method has converged, we calculate several entropies associated with the $\langle L_{\mathbb{I}}|$ and $|R_{\mathcal{O}}\rangle$ vectors: most notably, the Von Neumann entropy $S_t^{VN}(L_{\mathbb{I}})$ and the generalized Rényi 2, which we define as $S_t^{r2}(R_{\mathcal{O}}) = -\ln \sum_n \lambda_n^2$, where λ_n^2 are the eigenvalues of the
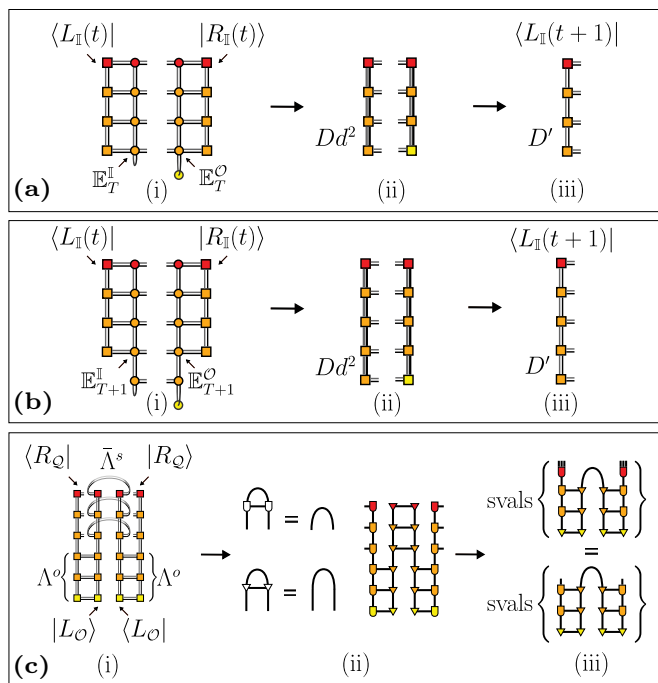
FIG. 5. Illustrations of the methods used for building the temporal matrix product states (tMPS). (a) Power method. (b) Light cone method. (c) The matrix involved in our low-rank approximation of $\mathcal{T}$ and the role of the gauge transformation in our calculation.

FIG. 6. Fidelity for overlap $\langle L_{\mathcal{O}}^{D_{\max}}|L_{\mathcal{O}}^{D}\rangle$ for different bond dimensions, as afunction of time, for the integrable case with starting state $|+\rangle$. After fixing a threshold ϵ, we can extract the required bond dimension to give a faithful approximation of $|L_{\mathcal{O}}^{D}\rangle$.

RTM $\mathcal{T}_t^{L_{\mathcal{O}}|R_{\mathcal{O}}}$ [49]. Convergence is reached at a given step i of the power method if, for the aforementioned entropies, one has $\sum_{t=0}^{N_T}(S_t^{i-1} - S_t^i)^2 < \epsilon$, which in our case we take to be $\epsilon = 10^{-6}$.

The *light-cone method* works in a slightly different way, as it allows us to build the tMPS for a time T starting from the one for $T-1$ with a single optimization [see Fig. 5(b)]: Starting from $\langle L_{\mathbb{I}}|$ and $|R_{\mathbb{I}}\rangle$ of a given length N_T, we apply transfer matrices $\mathbb{E}^{T+1}$ of length $N_T + 1$ to the left and $\mathbb{E}_{\mathcal{O}}^T$ to the right, extending the network inboth the time and space directions.

As for the power method, after each iteration, the bond dimension of the tMPS grows by a factor d^2, so that truncating is required to avoid an exponential computational cost. The truncated $\langle L_{\mathbb{I}}|$ and $|R_{\mathbb{I}}\rangle$ are then used as starting points for the next iteration. The cone is thus built from the center moving outward, assuming that the truncation of its edges does not spoil the causal structure induced by the local operator. The algorithm seems in any case to be quite stable, as we checked by piling up periodically a few layers of $\mathbb{E}^T$ before truncating and verifying that the final result is the same as the one obtained when we truncate after each iteration.

Despite their differences, both methods end up giving comparable results, as they share the same truncation procedure, which goes as follows: We focus on optimizing the overlap between $\langle L_{\mathbb{I}}|$ and $|R_{\mathcal{O}}\rangle$, which includes the operator $\mathcal{O}$ whose expectation value we are interested in. Since our aim is to find a low-rank approximation for the RTM $\mathcal{T}_t^{L_{\mathbb{I}}|R_{\mathcal{O}}}$ at any bipartition t and $T-t$, we focus on computing the eigenvalues
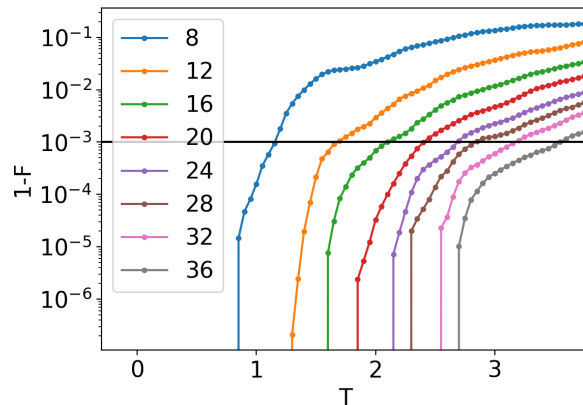
of the matrix $(\mathcal{T}_t^{L_{\mathbb{I}}|R_{\mathcal{O}}})^\dagger \mathcal{T}_t^{L_{\mathbb{I}}|R_{\mathcal{O}}}$ depicted in Fig. 5(c), which can be associated with the squares of the relevant singular values. As usual in TN calculations, we can transform a global optimization problem into a local one by making use of gauge transformation [50,51]. We start by bringing both L and R vectors individually to a standard orthogonal gauge, starting from the side of the initial state [Fig. 5(c)]. This gauge transformation allows-us to incorporate the $\bar{\Lambda}_{L,R}^s(t)$ factors toward the operator side, so that we do not need to compute them explicitly, as their contraction up to $t < T$ now reduces to an identity. We now only need to contract the bottom environments and project them over their largest D' singular values at each t, thus optimizing the overlap between the two tMPS. The advantage of keeping the local operators on either the left or right side of the network is that, while we optimize with respect to $\langle \mathcal{O} \rangle$, we always end up with a new $\langle L_{\mathbb{I}}|$ ($|R_{\mathbb{I}}\rangle$) which does not include the operator itself. This tMPS, which is related rather to the time evolution of the state, thus encodes the influence functional of the system and can be used as astarting point for the next iteration of our algorithms.
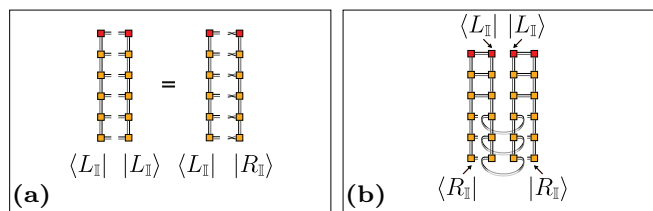

FIG. 7. (a) A possible method for truncating the temporal matrix product state (tMPS) relies on the optimization of the overlap $\langle L|L\rangle$, where in the symmetric case, $|L\rangle$ can be related by a partial transpose to the vector $|R\rangle$. (b) Alternatively, it is also possible to consider the reduced transition matrix (RTM) starting from the side of the initial state, a procedure which can give a nontrivial singular value decomposition (SVD) spectrum even in the absence of a local operator.
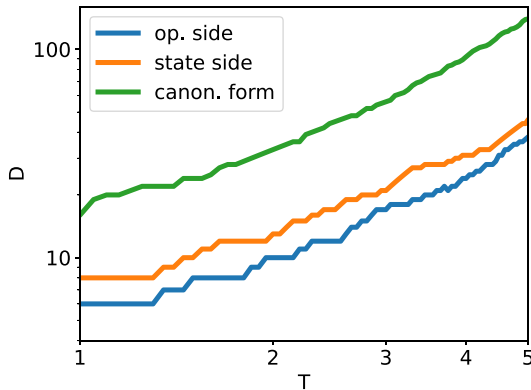
FIG. 8. Bond dimensions of the temporal matrix product state (tMPS) as a function of time obtained with different methods (as described in the text): our method (blue line), the optimization starting from the side of the initial state (orange), and the canonical form (green), imposing the same truncation error ($\epsilon_{\text{trunc}} = 10^{-6}$). Our method leads to the smallest bond dimension.

APPENDIX B: ESTIMATE OF THE BOND DIMENSION FOR A GIVEN FIDELITY

To estimate the bond dimension required to obtain our tMPS with a given fidelity, we follow a procedure like the one employed in Ref. [19] to estimate the truncation error in our tMPS compared with an exact result, which in our case corresponds to the evolution obtained with the maximum bond dimension we can afford $D_{\text{max}}(\simeq 1000)$. The largest T's we consider for this estimation are thus restricted to times for which D_{max} does not induce any sizable truncation error with respect to the exact (Trotterized) dynamics. If we work with normalized MPSs, the error induced by truncating the operator MPS to bond dimension D is encoded in the fidelity $F = |\langle \psi_O^{D_{\text{max}}}(t)|\psi_O^D(t)\rangle|$, where $|\psi_O^D(t)\rangle$ represents the time-evolved state up to time t truncated at every time step to bond dimension D. By fixing a given accuracy ϵ, we thus identify the maximum time for which $F \geqslant 1 - \epsilon$. As a consequence, by repeating the above procedure for different D, we can identify a curve $D^*(t)$ along which the truncation error is kept roughly constant at a value ϵ.

We then extract a similar curve $D^*(t)$ for the tMPS encoding the semi-infinite left (right) TN $\langle L_\mathcal{O}|$ ($|R_\mathcal{O}\rangle$). The truncation here is based on projecting onto the largest singular values of each RTM, as described above. The largest D_{max} we can afford is again used as the exact reference. The overlap of the two tMPSs measures the effects of the truncation error $F = |\langle L_\mathcal{O}^{D_{\text{max}}}|L_\mathcal{O}^D\rangle|$. Once more by fixing such an error to a given threshold ϵ, we can identify the corresponding $D^*(t)$

for the tMPS. The typical behavior of the resulting curves for F can be seen in Fig. 6.

APPENDIX C: COMPARISON WITH OTHER APPROACHES

As previously mentioned, our optimization prescription is a relatively small modification with respect to already existing algorithms, which allows us nevertheless to make a clearer connection with the underlying structures and estimate the computational complexity associated with the calculation of time-dependent expectation values. It might nevertheless be interesting to check how this method compares with other prescriptions employed in the literature.

The most straightforward optimization procedure when dealing with (t)MPSs after applying a column of MPOs is to truncate using the standard canonical forms for the left and right vectors separately. In our framework, with left-right symmetry of the Hamiltonian and translational invariance, this would amount to an optimization of $\langle L|$ with its conjugate $|L\rangle$ instead of $|R\rangle$. The operation of complex conjugation here can be seen as a partial transpose involving the forward and backward leg of the tMPS (basically an insertion of a series of swap operators), see Fig. 7(a), since (not considering additional complications in case the initial state is complex) it basically amounts to an exchange $U(t) \leftrightarrow U^\dagger(t)$ of the time evolution operators. Due to the nontrivial structure induced by this, the projector structure of the TN is lost even if no operators are present, so that one obtains a nontrivial bond dimension even when $\mathcal{O} = \mathbb{I}$.

Instead of optimizing with respect to a single operator inserted at the edge of the MPO column, the conjugation would imply acting on the whole column, i.e., to perform these transpositions at each timestep.

Another possible strategy, which has been suggested in Refs. [6,10], involves bringing the tMPS to canonical form starting from the bottom side and then doing a sweep from the initial state optimizing $\langle L_\mathbb{I}|R_\mathbb{I}\rangle$, Fig. 7(b). In this case, while the spectrum of the RTM is trivial since no operator is present, the SVD is not: This can be seen again as due to an insertion of swap operators which generate a nontrivial spectrum, which however cannot be directly related to the causal structure generated by a local operator. In this case, one is unable to give *a priori* a statement on the maximum rank required by the algorithm.

Despite this difference, remarkably, this latter method returns comparable bond dimensions with the ones obtained with our method when $\mathcal{O} = \sigma_x$, see Fig. 8. On the other hand, the optimization using the standard canonical forms requires a larger bond dimension, although the behavior of the various algorithms is comparable: This confirms our expectation that the scaling of our truncation based on the overlap $\langle L|R\rangle$ is the same as the one based on $\langle L|L\rangle$.

[1] P. Calabrese and J. Cardy, Evolution of entanglement entropy in one-dimensional systems, J. Stat. Mech.: Theory Exp. (2005) P04010.

[2] A. Laeuchli and C. Kollath, Spreading of correlations and entanglement after a quench in the one-dimensional Bose-Hubbard model, J. Stat. Mech.: Theory Exp. (2008) P05018.

[3] J. Dubail, Entanglement scaling of operators: A conformal field theory approach, with a glimpse of simulability of long-time dynamics in 1+1d, J. Phys. A: Math. Theor. **50**, 234001 (2017).

[4] M. C. Bañuls, M. B. Hastings, F. Verstraete, and J. I. Cirac, Matrix product states for dynamical simulation of infinite chains, Phys. Rev. Lett. **102**, 240603 (2009).

[5] A. Müller-Hermes, J. I. Cirac, and M. C. Bañuls, Tensor network techniques for the computation of dynamical observables in 1D quantum spin systems, New J. Phys. **14**, 075003 (2012).

[6] M. B. Hastings and R. Mahajan, Connecting entanglement in time and space: Improving the folding algorithm, Phys. Rev. A **91**, 032306 (2015).

[7] A. Strathearn, P. Kirton, D. Kilda, J. Keeling, and B. W. Lovett, Efficient non-Markovian quantum dynamics using time-evolving matrix product operators, Nat. Commun. **9**, 3322 (2018).

[8] C. D. White, M. Zaletel, R. S. K. Mong, and G. Refael, Quantum dynamics of thermalizing systems, Phys. Rev. B **97**, 035127 (2018).

[9] T. Rakovszky, C. W. von Keyserlingk, and F. Pollmann, Dissipation-assisted operator evolution method for capturing hydrodynamic transport, arXiv:2004.05177.

[10] M. Frías-Pérez and M. C. Bañuls, Light cone tensor network and time evolution, Phys. Rev. B **106**, 115117 (2022).

[11] A. Lerose, M. Sonner, and D. A. Abanin, Influence matrix approach to many-body Floquet dynamics, Phys. Rev. X **11**, 021040 (2021).

[12] A. Lerose, M. Sonner, and D. A. Abanin, Overcoming the entanglement barrier in quantum many-body dynamics via space-time duality, Phys. Rev. B **107**, L060305 (2023).

[13] G. Giudice, G. Giudici, M. Sonner, J. Thoenniss, A. Lerose, D. A. Abanin, and L. Piroli, Temporal entanglement, quasiparticles, and the role of interactions, Phys. Rev. Lett. **128**, 220401 (2022).

[14] J. Thoenniss, A. Lerose, and D. A. Abanin, Nonequilibrium quantum impurity problems via matrix-product states in the temporal domain, Phys. Rev. B **107**, 195101 (2023).

[15] E. Tirrito, N. J. Robinson, M. Lewenstein, S.-J. Ran, and L. Tagliacozzo, Characterizing the quantum field theory vacuum using temporal Matrix Product states, arXiv:1810.08050.

[16] R. Feynman and F. Vernon, The theory of a general quantum system interacting with a linear dissipative system, Ann. Phys. **24**, 118 (1963).

[17] M. Sonner, A. Lerose, and D. A. Abanin, Influence functional of many-body systems: Temporal entanglement and matrix-product state representation, Ann. Phys., **435**, 168677 (2021).

[18] E. Ye and G. K.-L. Chan, Constructing tensor network influence functionals for general quantum dynamics, J. Chem. Phys. **155**, 044104 (2021).

[19] T. Prosen and M. Žnidarič, Is the efficiency of classical simulations of quantum dynamics related to integrability? Phys. Rev. E **75**, 015202(R) (2007).

[20] I. Pižorn and T. Prosen, Operator space entanglement entropy in *XY* spin chains, Phys. Rev. B **79**, 184416 (2009).

[21] V. Alba, J. Dubail, and M. Medenjak, Operator entanglement in interacting integrable quantum systems: The case of the rule 54 chain, Phys. Rev. Lett. **122**, 250603 (2019).

[22] B. Bertini, P. Kos, and T. Prosen, Operator entanglement in local quantum circuits I: Chaotic dual-unitary circuits, SciPost Phys. **8**, 067 (2020).

[23] B. Bertini, P. Kos, and T. Prosen, Operator entanglement in local quantum circuits II: Solitons in chains of qubits, SciPost Phys. **8**, 068 (2020).

[24] L. Piroli, B. Bertini, J. I. Cirac, and T. Prosen, Exact dynamics in dual-unitary quantum circuits, Phys. Rev. B **101**, 094304 (2020).

[25] G. Vidal, Efficient simulation of one-dimensional quantum many-body systems, Phys. Rev. Lett. **93**, 040502 (2004).

[26] S. R. White and A. E. Feiguin, Real-time evolution using the density matrix renormalization group, Phys. Rev. Lett. **93**, 076401 (2004).

[27] S. Paeckel, T. Köhler, A. Swoboda, S. R. Manmana, U. Schollwöck, and C. Hubig, Time-evolution methods for matrix-product states, Ann. Phys. **411**, 167998 (2019).

[28] M. C. Bañuls, Tensor network algorithms: A route map, Annu. Rev. Condens. Matter Phys. **14**, 173 (2023).

[29] This can be easily generalized to an arbitrary number of sites if it is finite.

[30] I. P. McCulloch, From density-matrix renormalization group to matrix product states, J. Stat. Mech.: Theory Exp. (2007) P10014.

[31] B. Pirvu, V. Murg, J. I. Cirac, and F. Verstraete, Matrix product operator representations, New J. Phys. **12**, 025012 (2010).

[32] Notice that this definition differs slightly from those of previous works, since the specific operators $\mathcal{OQ}$ are included in the definition of the tMPS. By construction, if we chose as two-site operators the identity operator, $\mathcal{OQ} = \mathbb{I}$, we obtain that $\langle L_{\mathbb{I}} | R_{\mathbb{I}} \rangle = \langle \psi(t) | \psi(t) \rangle = 1$.

[33] Y. Nakata, T. Takayanagi, Y. Taki, K. Tamaoka, and Z. Wei, New holographic generalization of entanglement entropy, Phys. Rev. D **103**, 026005 (2021).

[34] K. Doi, J. Harper, A. Mollabashi, T. Takayanagi, and Y. Taki, Pseudo entropy in dS/CFT and time-like entanglement entropy, Phys. Rev. Lett. **130**, 031601 (2023).

[35] K. Narayan, de Sitter space, extremal surfaces and time-entanglement, Phys. Rev. D **107**, 126004 (2023).

[36] X. Jiang, P. Wang, H. Wu, and H. Yang, Timelike entanglement entropy and $T\bar{T}$ deformation, arXiv:2302.13872.

[37] For reflection-invariant Hamiltonians, $|L_{\mathcal{O}}\rangle$ differs from $|R_{\mathcal{O}}\rangle$ only by a complex conjugation.

[38] A. Coser, E. Tonni, and P. Calabrese, Entanglement negativity after a global quantum quench, J. Stat. Mech.: Theory Exp. (2014) P12017.

[39] P. Calabrese, J. Cardy, and E. Tonni, Entanglement negativity in quantum field theory, Phys. Rev. Lett. **109**, 130502 (2012).

[40] G. Vidal and R. F. Werner, Computable measure of entanglement, Phys. Rev. A **65**, 032314 (2002).

[41] P. Calabrese and J. Cardy, Entanglement and correlation functions following a local quench: A conformal field theory approach, J. Stat. Mech.: Theory Exp. (2007) P10004.

[42] F. Verstraete and J. I. Cirac, Continuous matrix product states for quantum fields, Phys. Rev. Lett. **104**, 190405 (2010).

[43] W. Tang, F. Verstraete, and J. Haegeman, Matrix product state fixed points of non-Hermitian transfer matrices, arXiv:2311.18733.

[44] X. Wang and T. Xiang, Transfer-matrix density-matrix renormalization-group theory for thermodynamics of one-dimensional quantum systems, Phys. Rev. B **56**, 5061 (1997).

[45] Y.-K. Huang, Biorthonormal transfer-matrix renormalization-group method for non-Hermitian matrices, Phys. Rev. E **83**, 036702 (2011).

[46] For operators with known small bond dimension such as σ_z or $\sigma_x\sigma_x$, our algorithm saturates at the expected fixed value of D [47].

[47] M. J. Hartmann, J. Prior, S. R. Clark, and M. B. Plenio, Density matrix renormalization group in the Heisenberg picture, Phys. Rev. Lett. **102**, 057202 (2009).

[48] A. Foligno, T. Zhou, and B. Bertini, Temporal entanglement in chaotic quantum circuits, Phys. Rev. X **13**, 041008 (2023).

[49] This choice ensures that the RTM is symmetric, which guarantees that its diagonalization is not an ill-posed problem.

[50] R. Orús and G. Vidal, Infinite time-evolving block decimation algorithm beyond unitary evolution, Phys. Rev. B **78**, 155117 (2008).

[51] G. Evenbly, A practical guide to the numerical implementation of tensor networks I: Contractions, decompositions and gauge freedom, arXiv:2202.02138.