

Measurement and Entanglement Phase Transitions in All-To-All Quantum Circuits, on Quantum Trees, and in Landau-Ginsburg Theory

Adam Nahum^{1,*}, Sthitadhi Roy^{1,2}, Brian Skinner³ and Jonathan Ruhman^{4,5,†}

¹*Rudolf Peierls Centre for Theoretical Physics, Clarendon Laboratory, Oxford University, Parks Road, Oxford OX1 3PU, United Kingdom*

²*Physical and Theoretical Chemistry, Oxford University, South Parks Road, Oxford OX1 3QZ, United Kingdom*

³*Department of Physics, Ohio State University, Columbus, Ohio 43210, USA*

⁴*Department of Physics, Bar-Ilan University, Ramat Gan 52900, Israel*

⁵*Center for Quantum Entanglement Science and Technology, Bar-Ilan University, Ramat Gan 52900, Israel*



(Received 10 October 2020; accepted 19 January 2021; published 30 March 2021)

A quantum many-body system whose dynamics includes local measurements at a nonzero rate can be in distinct dynamical phases, with differing entanglement properties. We introduce theoretical approaches to measurement-induced phase transitions (MPTs) and also to entanglement transitions in random tensor networks. Many of our results are for “all-to-all” quantum circuits with unitaries and measurements, in which any qubit can couple to any other, and related settings where some of the complications of low-dimensional models are reduced. We also propose field-theory descriptions for spatially local systems of any finite dimensionality. To build intuition, we first solve the simplest “minimal cut” toy model for entanglement dynamics in all-to-all circuits, finding scaling forms and exponents within this approximation. We then show that certain all-to-all measurement circuits allow exact results by exploiting local treelike structure in the circuit geometry. For this reason, we make a detour to give general universal results for entanglement phase transitions in a class of random tree tensor networks with bond dimension 2, making a connection with the classical theory of directed polymers on a tree. We then compare these results with numerics in all-to-all circuits, both for the MPT and for the simpler “forced-measurement phase transition” (FMPT). We characterize the two different phases in all-to-all circuits using observables that are sensitive to the amount of information that is propagated between the initial and final time. We demonstrate signatures of the two phases that can be understood from simple models. Finally we propose Landau-Ginsburg-Wilson-like field theories for the measurement phase transition, the forced-measurement phase transition, and for entanglement transitions in random tensor networks. This analysis shows a surprising difference between the measurement phase transition and the other cases. We discuss variants of the measurement problem with additional structure (for example free-fermion structure), and questions for the future.

DOI: [10.1103/PRXQuantum.2.010352](https://doi.org/10.1103/PRXQuantum.2.010352)

I. INTRODUCTION

A quantum system whose unitary dynamics is interspersed with repeated measurements follows a random trajectory through Hilbert space [1–5], determined both by the unitary part of the dynamics and by the sequence of measurement outcomes. In the many-body case this

random dynamics admits a “measurement phase transition” (MPT) between two qualitatively different, stable dynamical phases, with distinct entanglement properties [6–27]. For definiteness, consider a system of many spins in a pure state, evolving under a quantum circuit that includes both entangling two-spin unitary gates and measurements, which are made at random times at a finite rate per spin. Informally, sufficiently frequent measurements yield a “disentangling” phase: in this phase, the state at a given time is weakly entangled, and is fully specified by the outcomes of a relatively recent set of measurements. (The limiting case of this disentangling dynamics is where all the spins are measured simultaneously, leaving the system in a product state that can be read off from the measurement outcomes.) But when the frequency of measurements falls below a critical threshold,

*adam.nahum@physics.ox.ac.uk

†jonathan.ruhman@biu.ac.il

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/) license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

the dynamics enters an entangling phase [6,7]. In this phase the dynamics produces states with extensive entanglement, which retain quantum information from much earlier times. If the initial state is mixed, rather than pure, then it will rapidly be purified [12] by the repeated measurements in the disentangling phase, but not in the entangling phase.

The simplest toy model for the MPT arises from thinking about the connectivity of the space-time diagram of the quantum circuit, viewed as a tensor network [6]. In this representation a measurement event is a break in the worldline of a spin, across which quantum information cannot be transmitted. When measurements become sufficiently frequent, the circuit falls apart into disconnected pieces, implying that entanglement in the final state is short ranged and there is no transmission of quantum information, from the initial to the final state, over long timescales.

The existence of the MPT poses several types of questions. Viewing the circuit as a quantum information processor, the MPT is a transition in the properties of a randomly generated error-correcting code [11,24,28], the structure of which can be optimal in a certain sense [12]; understanding the MPT may lead to useful insights into fault-tolerant quantum computation [29].

The transition also has consequences for the computational difficulty, for a *classical* computer, of simulating various types of open or monitored quantum systems [6,30–33]. As a simplified thought experiment, we may imagine that we are given the sequence of measurement outcomes obtained in an experimental run (as well as the information about the Hamiltonian, and the initial state), and asked to calculate the quantum state of the system following these measurements. If the dynamics is in the disentangling phase, an efficient matrix-product or tensor-network representation of the evolving state may be possible, while if the dynamics is in the entangling phase, the computation may be intractable. In this sense the MPT can function as an “epistemological” phase transition in which the quantum wave function becomes essentially unknowable.

Philosophically, we may also wonder what the existence of two phases implies about how to distinguish dynamical processes that are intrinsically quantum from those that are effectively classical. For example, in both of the phases separated by the MPT, quantum correlations between local observables are “weak,” but for different reasons. In the disentangled phase, a local operator is correlated only with a few others nearby. In the entangled phase, it may have nontrivial correlations, but these are detectable only by highly nonlocal “scrambled” operators, and hidden from local ones. Only close to the transition point does the system escape both mechanisms, allowing nontrivial correlations for local operators [6,9,13–17,24]. Yet another key question is how to probe the MPT experimentally [13,17].

This question is nontrivial: for example, a naive approach leads to a severe sampling problem (due to the need to compare measurements in distinct experimental runs that have the same measurement outcomes).

Another way of looking at the MPT is as a problem in statistical mechanics and critical phenomena [6,7,9,10,12–16,22,24,25,27,34]. Open questions abound, both about the nature of the phases and about the critical point separating them. Many variants of the measurement transition can be imagined; how do we sort them into universality classes? Are there simplifying limits where exact results are possible? Are there useful continuum field theories for the MPT and related problems, that allow us to apply the tools of the renormalization group?

This statistical mechanics problem is closely connected to an entanglement transition that takes place in random tensor networks [14,35,36] (we explore the similarities and differences further here) and the same questions apply in that setting. These problems are challenging partly because of the need to average over randomness: either intrinsic randomness in the definition of the dynamics (for example, if we consider dynamics using a random quantum circuit) or simply the inevitable quantum-mechanical randomness in measurement outcomes.

Our focus in this paper is on circuits built from generic unitary gates (for example, Haar-random gates). An alternative profitable direction is to study circuits made from Clifford unitaries [7,12,15,17,22,37–41]. Clifford circuits are efficiently classically simulable, which has allowed direct tests of conformal invariance at the MPT in $1 + 1D$ [9,15] and simulations in $2 + 1D$ [22]. In general, the universality class of the MPT is expected to differ for Clifford versus generic unitaries (see, e.g., Ref. [16]), though many features of the stable phases are similar.

This paper is a journey through several approaches to the MPT, and also to the closely related “forced” measurement phase transition (FMPT, defined in Sec. II A below), and to the entanglement transition in various types of random tensor networks (RTNs). Our aim is to find settings in which exact results can be obtained for the transition, as well as to clarify the properties of the two phases. We examine several different tools and settings, but the unifying feature is that we consider measurement and entanglement transitions in situations where the complications arising from low-dimensional spatial structure are reduced.

Much of this paper is concerned with circuits with all-to-all couplings between qubits, i.e., with no fixed spatial geometry, which we study using analytical arguments and numeric simulations. (Various types of all-to-all circuit have also been discussed recently in Refs. [12,26].) These circuits are in turn closely related to tree tensor networks, for which we give exact results, including the first exact identification of an entanglement transition in a generic system with finite bond dimension.

Turning to models with a finite number of spatial dimensions, we discuss and extend tools based on mappings to effective “lattice magnets” [13,14,34–36,42–45], involving a replica limit [13,14,34,36], which capture the properties of the two phases, and in principle the critical point. We suggest alternative ways of thinking about these effective models, making connections with ideas from disordered magnetism: in particular, we suggest a construction of order parameters for the MPT and for entanglement transitions in random tensor networks, based on overlap of Feynman trajectories. A key outstanding question is the existence of effective field theories for the MPT. Here we propose—speculatively—two Landau-Ginsburg theories, one for the MPT and one for both the FMPT and the RTN.

The sketch in Fig. 1 contrasts an all-to-all measurement circuit and a circuit with a fixed spatial geometry. In this figure, time runs vertically, and each worldline represents a spin or qubit. Unitary gates are applied between randomly chosen spins at random times, and projective measurements are applied to randomly chosen spins. All-to-all coupling is perhaps the simplest setting for the MPT. Since the distinction between area and volume law breaks down in the all-to-all case (as also in the limit of infinite dimensions), it is natural to focus instead on the transmission of information between initial and final times. Here we characterize this transmission via the operator entanglement [46–51] of the nonunitary time-evolution operator, defined below. This quantity has a simple interpretation in terms of the surface tension of the “entanglement membrane” in the effective replica description, which we discuss. An even simpler heuristic picture for it comes from the classical toy model, in terms of the minimal cut that separates the top of the circuit from the bottom.

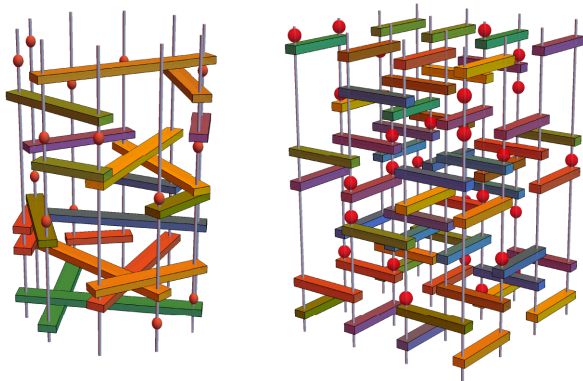


FIG. 1. Random circuits with measurements. Vertical lines are world lines of individual spins, with time running vertically. The blocks connecting different world lines are independently random unitary gates. These are interspersed by projective measurements, represented by the red blobs. The left panel shows an all-to-all two-local circuit whereas the right panel shows a regular 2 + 1D circuit.

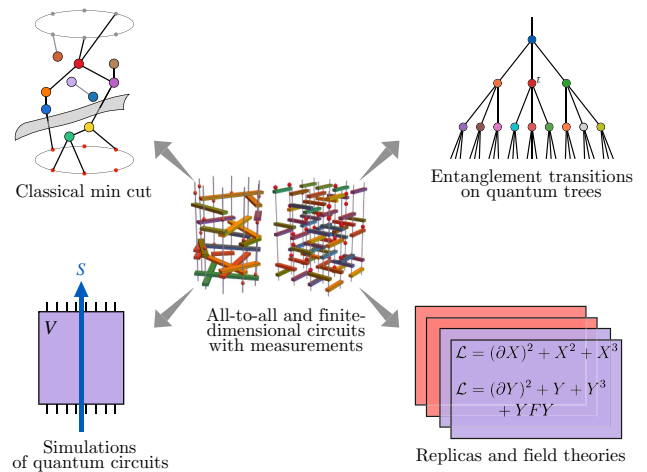


FIG. 2. Some of the approaches in this paper. To gain intuition, we start by solving a classical toy model for the transition in the all-to-all circuit, which gives scaling forms for the “minimal-cut” cost that determines the zeroth Rényi entropy (Sec. III). In Sec. IV we turn to the truly “quantum” problem, obtaining exact results for random tree tensor networks. These can be applied to the true quantum transition in all-to-all circuits with “forced” measurements. Section V simulates all-to-all measurement circuits, using operator entanglement of the time-evolution operator, and the convergence with time of two initially orthogonal states, to diagnose the preservation of information over an exponentially long timescale that is a hallmark of the entangling phase. Section VI develops analytical approaches to the MPT and to entanglement transitions based on the replica trick, clarifying the properties of the two phases and suggesting candidate field theories for the critical points in various settings.

We apply all the approaches mentioned above (tree approximations, simulations, replica field theories) in the setting of generic quantum circuits for spin 1/2, as well as related random tensor networks, giving results for scaling properties in the entangled phase and close to the critical point. We also study a solvable “classical” limit of the problem. Our main approaches are illustrated in Fig. 2, and the ensuing section, Sec. II, gives an overview of our results. In closing this Introduction, however, let us briefly clarify the logic of our four-pronged approach to understanding the MPT and its relatives.

Before tackling the “true” quantum circuit problem, we find it instructive to first solve the classical toy model mentioned above, in the particular setting of all-to-all circuits (Sec. III). In this model the entanglement is described in terms of a “minimal cut” through a circuit in which worldlines have been broken by measurement. The minimal cut becomes an exact description of the MPT in certain limits, but in general it does *not* capture either the location of the critical point or the true critical scaling of the quantum problem. Nonetheless, the minimal cut problem yields some useful lessons for the full quantum problem. Most prominently, it captures key qualitative

features of the two phases, including the appearance of an exponentially long timescale for survival of quantum information within the entangled phase. Solving the minimal cut problem also makes clear certain crucial concepts for understanding the MPT in all-to-all circuits, including the local tree structure of the circuit and the relevance of crossover scaling phenomena.

The fact that all-to-all circuits have a local tree structure then motivates us to study entanglement transitions in quantum trees (Sec. IV). In this setting we are able to obtain the exact location of the entanglement transition (and exact critical properties) for a tree tensor network that is relevant to dynamics with Haar-random gates. This result may be useful for further investigations: studies of the MPT in systems with generic unitaries are often hampered by the restriction of numerics to small sizes, which make it difficult to accurately pinpoint entanglement transitions. Moreover, we argue that the critical measurement rate that we identify in the quantum tree is also the exact result for the full all-to-all quantum circuit with forced (postselected) measurements.

Armed with the understanding gained from the minimal-cut and quantum tree problems, we turn our attention to direct numerical simulations of the quantum circuit (Sec. V). The results we obtain are consistent with the critical scaling forms suggested by the previous approaches, and highlight the emergence of an exponentially long timescale associated with information transmission through the circuit in the entangling phase.

Finally, in Sec. VI we discuss mappings of the MPT and of random tensor networks to effective lattice models for a “pairing field,” and we discuss how to coarse grain such models. We construct the simplest candidate Lagrangians that are consistent with the replica symmetry and describe some of their features. We also touch on free fermions subject to measurement [52], which do not show the same kind of transition between weakly and strongly entangled phases but do show transitions of a different type [37,53,54]. We contrast these free systems (which have a continuous, rather than discrete, replica symmetry) with generic models, and we discuss some other variants of the MPT.

II. OVERVIEW

A. Models

Our starting point is a dynamical process in which a large number N of spin $1/2$ s undergoes unitary evolution punctuated by projective single-spin measurements: Fig. 1, left. (Circuits with both unitaries and measurements are referred to as “monitored” or “hybrid” quantum circuits.) The spins are “all-to-all” coupled, meaning that unitary gates may be applied between any two spins in the system. These gates are applied at a uniform rate between randomly chosen pairs of spins, and are themselves drawn

independently from a random ensemble (e.g., the Haar ensemble). Measurements, which are made in the Z basis, are also applied at a uniform rate to randomly chosen spins. The only parameter is $r \in [0, 1]$, which determines the relative rate of measurements and unitaries: in a unit interval of time there are on average rN measurements and $(1 - r)N$ unitary operations.

We distinguish between two possibilities for the projective measurements, which we refer to as “measurements” and “forced measurements,” respectively. (Correspondingly we refer to the “measurement phase transition,” or MPT, and “forced-measurement phase transition,” or FMPT.) The outcomes of “measurements” are determined as usual by the Born rule, based on the state of the system at the time of measurement. By contrast the probability of a given outcome for a “forced measurement” is independent of the state. We take it to be $1/2$ for both of the two possible outcomes, \uparrow and \downarrow —but in fact, for the ensembles of random unitaries we consider, it is completely equivalent to take all the measurement outcomes to be \uparrow . We can think of the FMPT as pertaining to a protocol in which we run (exponentially) many samples, discarding all those except those that yield the desired (“postselected”) sequence of outcomes.

To formalize the distinction between MPT and FMPT, define $V_{\mathbf{m}}$ to be the nonunitary time evolution operator represented by a given realization of the circuit. This operator is the product of unitaries and projection operators: we label it by a given sequence \mathbf{m} of outcomes for the measurement events: for example $\mathbf{m} = (\uparrow, \downarrow, \dots, \uparrow)$. ($V_{\mathbf{m}}$ also depends on the total time t , locations and times of the unitaries and measurements, and the specific random unitaries in the circuit realization, but we leave these dependencies implicit.) For the MPT, and for a given sequence of unitaries and measurement locations, the probability of a sequence of measurement *outcomes* \mathbf{m} is

$$P(\mathbf{m}) = \langle \psi(0) | V_{\mathbf{m}}^\dagger V_{\mathbf{m}} | \psi(0) \rangle, \quad (1)$$

where $|\psi(0)\rangle$ is the initial state. For the FMPT it is

$$P(\mathbf{m}) = 2^{-|\mathbf{m}|}, \quad (2)$$

where $|\mathbf{m}|$ is the number of measurements in a given realization of the circuit. In both cases, the time evolution of a pure state is

$$|\psi(t)\rangle = \frac{V_{\mathbf{m}} |\psi(0)\rangle}{|V_{\mathbf{m}} |\psi(0)\rangle|}. \quad (3)$$

It is occasionally useful to generalize the circuit to a variable number of spin states q for each site. In particular, the limit of large q is one way to motivate the classical problem we describe below.

Having started with the models above, we are led to consider some other related problems. These models are

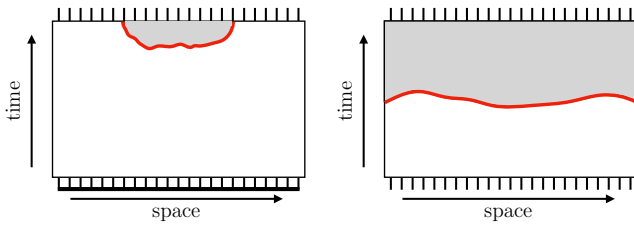


FIG. 3. The entanglement entropy of states and operators can be described in terms of the surface tension of an effective membrane or domain wall. Left: interpretation of entanglement entropy of a spatial subregion, in the entangling phase, as the free energy of an anchored membrane. Right: operator entanglement of the nonunitary circuit (in the entangling phase) in terms of a “horizontal” membrane (see also Ref. [13]).

introduced as we need them. Section IV considers a class of tree tensor networks, one example of which is closely related to the FMPT case above. Section VI addresses both circuits and tensor network models in a finite number of dimensions, in which we do have a sense of spatial locality.

B. Detecting the entangling phase

Before turning to the critical properties, we discuss the more basic issue of how to distinguish the two phases.

The entanglement transition can be identified with the vanishing of an effective *surface tension* for a membrane-like object in space time, as we discuss below. In the classical toy model, this membrane is a minimal cut through the circuit [6]. In a more precise picture, it is a domain wall in an effective statistical mechanics problem (see following sections). The surface tension of this membrane or domain wall is positive in the entangling phase [6].

In finite dimensions, the vanishing of this surface tension, which we denote $s_n(r)$ [55], implies a vanishing of the entanglement entropy density of the *states* produced by the dynamics at late time. This density is the coefficient of the volume law for the entanglement entropy of a spatial subregion, and is given by the surface tension $s_n(r)$. This is because the subregion’s entropy maps to the free energy of a membrane that is anchored on the boundary of region A on the final time surface: see Fig. 3 for a schematic in $1 + 1D$.

In the all-to-all circuit there is no distinction between areas and volumes (as in the limit of high dimensions), so the naive attempt to define an entropy density using the entropy of a spatial subregion is contaminated by trivial short-range entanglement (i.e. entanglement that could be removed with a shallow-depth circuit). Instead it is simpler to consider the entanglement properties of the *operator* that implements the time evolution itself. The operator entanglement [46–50], defined below, is a measure of the amount of quantum information transmitted from the initial to the final time by the nonunitary evolution operator V_m . In the membrane picture, this operator entanglement

is equal to the free energy of a “horizontal” membrane that completely traverses the system [13], as shown in Fig. 3. This observable also detects the vanishing surface tension $s_n(r)$ for the domain wall, as detailed below, but it does not require us to specify a spatial subregion.

Gullans and Huse proposed in Ref. [12] to think about dynamics with measurements in terms of the entropy of a state that starts out as maximally mixed, and is gradually purified by the dynamics. (The entangling phase is then a “mixed” phase, where the state remains mixed for a long time, and the disentangling phase is a “pure” phase where the state is rapidly purified.) This mixed state entropy is in fact equal to the operator entanglement of the nonunitary evolution operator. Reference [12] noted the exponentially long timescale for the survival of quantum information in the entangled phase (and plateaus in various observables), which will play an important role below. See also the recent Refs. [27,56].

Formally, the n th operator entanglement entropy of the circuit, denoted S_n throughout this paper, may be defined via the singular value decomposition of the nonunitary time-evolution operator V_m :

$$V_m \propto \sum_{j=1}^{2^N} \lambda_j |j\rangle_t \langle j|_0, \quad (4)$$

where $\{|j\rangle_0\}$ and $\{|j\rangle_t\}$ are bases corresponding to the initial and final time. Normalizing the λ_j so $\sum \lambda_j^2 = 1$,

$$S_n = \frac{1}{1-n} \ln \left(\sum_j \lambda_j^{2n} \right). \quad (5)$$

For the unitary case ($r = 0$), $S_n = N \ln 2$ is maximal at all times. For positive r , and for asymptotically late times, a single term dominates Eq. (4), meaning essentially that all initial states are projected onto the same final state—i.e., the final state can be read off from measurement outcomes \mathbf{m} (and the structure of the circuit) without knowledge of which initial state was fed in.

We also discuss another observable for quantifying the transmission of quantum information from initial to final times, which is more numerically tractable: this is the overlap between two initially orthogonal states, both subjected to the same V_m . (In the entangling phase, initially orthogonal states remain orthogonal for a long time.)

We characterize the operator entanglement in the classical toy model (Sec. III), in numerical simulations (Sec. V), using the replica trick (Sec. I), and with a toy model based on multiplying random matrices (Appendix E). The following basic points hold in all of these approaches.

First, in the entangling phase a nonzero density can be associated with the operator entanglement:

$$s_n(r) \equiv \lim_{t \rightarrow \infty} \lim_{N \rightarrow \infty} \frac{S_n(r, N, t)}{N}. \quad (6)$$

We think of this quantity as the information transmitted per spin, or in the membrane picture as the surface tension for a “horizontal” membrane. $s_n(r)$ is positive in the entangling phase, and vanishes continuously, for all $n \geq 1$, as the critical measurement rate r_c is approached from below.

As for almost any product of many random matrices, we expect that if N and r are fixed, then at sufficiently late times one of the singular values dominates the others and S_n decays exponentially in time. But if $s_n(r)$ is positive, this exponential decay does not set in until a time $\tau(r, N)$ that is exponentially large in N . We may define

$$a(r) = \lim_{N \rightarrow \infty} \frac{\ln \tau(r, N)}{N}. \quad (7)$$

Close to the transition, at $r \lesssim r_c$ where $s_n(r)$ is small,

$$a(r) \sim s_n(r) \quad (8)$$

(up to an order-1 constant of proportionality). On times t satisfying $\ln t \ll \ln \tau$, the entanglement deviates logarithmically from the “plateau” value dictated by Eq. (6). For example in one regime,

$$S_n(r, N, t) \simeq s_n(r)N - \ln t. \quad (9)$$

This formula has also been obtained in various limits in Refs. [12,27]: in particular Li and Fisher in Ref. [27] give a discussion very similar to that in Sec. I, in terms of domain walls in an effective quasi-1D model. In this interpretation $\ln t$ is the translational entropy of a domain wall. More generally, randomness and other effects can modify the nature of the subleading term above slightly, depending on the time regime.

In contrast to the above, the information transmitted per spin, $\lim_{N \rightarrow \infty} S_n(r, N, t)/N$, decays exponentially with t in the disentangled phase.

We now give an overview of our approaches to critical properties of these circuits and related models, considering each approach in turn (summarized in Fig. 2). The reader may obtain the key points of each approach from the corresponding Overview section. We also highlight some points that are not yet resolved, and places where our arguments rely on conjectures that could be tested further.

C. Min-cut toy model

Before attempting an exact treatment of the true quantum transition in the spin-1/2 circuit, we consider a limit (we sometimes refer to this as the “classical” limit) in

which the entanglement transition becomes a simple geometric problem involving a random graph. This graph is defined such that its edges represent the time evolution of each spin, which can be severed by measurement, and its nodes represent interactions (applied gates) between spins. The analog of the operator entanglement entropy is the cost of a “minimal cut” that disconnects the initial-time and final-time nodes: see Sec. III for a detailed definition.

Determining the scaling of this min-cut cost is a toy problem that provides intuition for the generic “quantum” problem. The minimal cut becomes an exact description of the operator entanglement only in special limits, as described in Sec. III (specifically, for projective measurements in the case where the local Hilbert space dimension q goes to infinity, and for a generic local Hilbert space dimension if we consider the somewhat unphysical zeroth Rényi entropy, S_0). The “classical” problem has its transition at a measurement rate r_c^{cl} that is, for spin 1/2, strictly larger than the critical measurement rate r_c for the true quantum transition, as diagnosed, for example, by all the S_n with $n \geq 1$.

We first identify the critical point r_c^{cl} associated with percolation on the graph, which illustrates the importance of local tree structure in all-to-all circuits. We then present an effective continuum field theory for percolation on this graph, which gives the relevant scaling forms near r_c^{cl} . We demonstrate this critical scaling using extensive numerical simulations for percolation observables and correlation functions. This demonstration is possible despite significant finite-time corrections, which arise because the critical timescale scales as $N^{1/5}$ and is modest even for simulations with very large N .

We demonstrate the plateau in the cost of the minimal cut that was described above, $S_0 \sim s_0(r)N$ over a long timescale. Close to criticality at $r \lesssim r_c^{\text{cl}}$ we find the entanglement density (min-cut tension)

$$s_0(r) \sim (r_c^{\text{cl}} - r)^{5/2}, \quad (10)$$

which is an appropriate limit of a general scaling form $S_0 = H(t/N^{1/5}, \delta r N^{2/5})$, and the corresponding long timescale

$$\tau \sim \exp[a(r) \times N], \quad a(r) \sim (r_c^{\text{cl}} - r)^{5/2}. \quad (11)$$

The scaling we identify applies not only for the all-to-all problem, but also for spatially local circuits with spatial dimension $d \geq 5$, as follows from a standard crossover scaling argument.

The all-to-all percolation model has also been analyzed independently in Ref. [57], using a different method in which rate equations for the number of percolation clusters of a given size are solved. This analysis also implies that the scaling variables are $t/N^{1/5}$ and $\delta r N^{2/5}$, in agreement with what we find.

D. Tree tensor networks: exact results

When the system size N is large, the structure of the quantum circuit in the vicinity of a given unitary is treelike (the smallest loops involve a parametrically large number of unitaries). This means that it is trivial to locate the classical critical point mentioned above. But in some cases (forced measurements) it also allows exact results for the quantum problem. This motivates us to study entanglement transitions on “quantum trees,” i.e., tree tensor networks, in Sec. IV.

While our approach could be generalized, we focus on trees with bond dimension 2, where each node is a random tensor whose probability distribution is invariant under $U(2)$ rotations on its legs. This includes trees that appear spontaneously in the spin-1/2 FMPT circuit for unitaries drawn from the Haar measure, for example.

Formally we can think of an (upside-down) tree like that shown in Fig. 2 as a tensor-network wave function for a single spin at the apex (root) and many spins (leaves) at the base. Our starting point is to characterize the entanglement between apex and base, which for a bond dimension 2 tree is characterized by a single number, Z . For an asymptotically large tree, Z has a critical vanishing at a particular measurement rate r_c . (In more general trees, r can be thought of as a parameter in the node tensors’ distribution.)

We write a random recursion relation for Z as a function of the generation number k of the tree. This recursion relation allows us to derive the location of the critical point r_c analytically for the case with Haar-random unitaries (we also study a slightly broader class of distributions):

$$r_c = \frac{212 + 75\pi}{362 + 75\pi}. \quad (12)$$

This critical point r_c is detected by any Rényi entropy S_n with $n > 0$; S_0 instead detects the classical transition, at the strictly larger value r_c^{cl} , discussed in the previous section. This is the difference between the existence of a percolating path connecting the root of the tree to infinity (for $r < r_c^{\text{cl}}$) and the ability of the tree to broadcast a nonzero amount of quantum information from the root of the tree to infinity, rather than an amount that decays exponentially with the distance from the root.

Assuming a plausible conjecture, Eq. (12) is also the value of r_c for the FMPT in the all-to-all Haar circuit, and yields a bound on the critical scaling of the entanglement density $s_2(r)$ [Eq. (6)]. While the treatment of the tree may hold lessons for the MPT in addition to the FMPT, we do not discuss the MPT from this perspective: the measurement correlations encoded in Eq. (1) hamper our approach.

We also obtain the critical scaling of Z for $r \lesssim r_c$. Since the full nonlinear recursion relation for Z is complicated, this requires us to make a *conjecture*, which is that the universal features of the scaling are faithfully retained

in a simplified nonlinear recursion relation. We can then write a continuum description that describes a Fisher-Kolmogorov-Petrovsky-Piskunov- (FKPP) like traveling wave [58]. This is a standard description for the partition function of a directed polymer on a tree [58], with the addition of a diffusion constant that varies with the (fictitious) spatial coordinate, reflecting the nonlinearity of the recursion. For the parameters of the trees we treat, there is a surprisingly rapid scaling close to the critical point: the entanglement between apex and base of an infinite tree scales as ($S_2^{\text{tree}} \simeq 2Z$)

$$S_2^{\text{tree}} \sim \exp\left(-\frac{\text{const}}{\sqrt{r_c - r}}\right). \quad (13)$$

We also address the scaling of S_n as a function of tree size exactly at r_c .

Using a nonrigorous argument, we extend these formulas to give the entanglement of a subset of the spins, in a tree tensor-network wave function for a spin chain (whose spins are the leaves of the tree). These results are not relevant to the all-to-all circuit, but are interesting in the context of tree tensor-network states, which are toy models for some features of scale invariance in $1 + 1\text{D}$, and are also useful numerical tools [59–68]. We obtain a “modified minimal cut” formula for the tree, in which the cost of cutting a bond in the tree is loosely speaking weighted by appropriate factors of the quantity Z , which is parametrically small close to r_c . This gives a quantitative picture of how the entanglement of ℓ consecutive spins in a tree tensor-network state goes from the well-known logarithmic scaling, $S_\ell \sim c(r) \ln \ell$, suggested by its hierarchical structure [67–69], to an area law state, $S_\ell = \mathcal{O}(\ell^0)$. We find that $c(r)$ vanishes exponentially as $r \rightarrow r_c$ and that the state is area law even at r_c .

Recently, Ref. [68] studied the entanglement transition in a quantum tree state, with bond dimension 3, using a different approach. The authors conjectured that the scaling was the same as in a statistical mechanics model that shares some of the features of a replica formulation derived from the tree (the exact replica formulation was not tractable). Surprisingly, the findings in Ref. [68] are quite different from those we obtain (assuming the conjecture mentioned in the previous paragraph) in the trees studied here. For example, Ref. [68] finds that the coefficient c in $S \sim c \ln \ell$ is a power law in the control variable close to the transition, and that entanglement is super-area-law at r_c . The reason for the different results in these two models remains to be understood.

Our conjectured continuum description for scaling in the tree has a parameter Δ that describes the degree of disorder in the tensor network, and which determines the critical exponents. For the trees we study, whose node tensors have a distribution with a $U(2)$ invariance property on the legs, this parameter is fixed to $\Delta = 1/4$ at r_c . This corresponds

to a “strong disorder” regime [58]. We raise the question of whether general distributions of tensors allow us to explore the phase transition at other values of Δ . If so, it is possible to obtain a range of universality classes for the tree transition, analogous to a renormalization group fixed line. However, we do not determine whether this is possible.

E. Direct simulations of quantum circuits

We perform direct simulations of the all-to-all measurement circuit and forced-measurement circuit, and interpret the results in the light of the tree calculation and the replica approach described below. These simulations are computationally demanding: we are limited to system sizes $N \leq 20$ for quantities involving states and to smaller sizes for the operator entanglement. Determining r_c accurately (the value of which is expected to differ for measurements and forced measurements) is not possible, but we are able to confirm many of the key features of the entangled phase in Sec. II B.

We give evidence for the plateau, Eq. (6), in the operator entanglement, with a nonzero information transmission per spin $s(r)$ that is asymptotically time independent, and for a positive exponential growth coefficient $a(r) > 0$ for the characteristic timescale within the entangled phase.

It is convenient to define this timescale τ via the late-time convergence of two distinct, initially orthogonal, states $|\psi_1(t)\rangle$ and $|\psi_2(t)\rangle$ that are postselected to undergo the *same* sequence of measurement outcomes, so that they are evolved with the same $V_{\mathbf{m}}$. These states remain approximately orthogonal for a long time in the entangled phase: a kind of effective unitarity of the nonlinear, nonunitary time evolution Eq. (3) for a given \mathbf{m} . (This orthogonality is related to the error-correction property of the dynamics [11,12].) The two states collapse at late times. We show that $a(r)$ is positive at small r and vanishes at large r .

For forced measurements our expectation is that r_c is given by the result of the tree calculation. Numerically, it in fact becomes unmeasurably small at a significantly smaller value of r . Our interpretation of this is that, because of exponential scaling in Eq. (13), the quantities $s(r)$ and $a(r)$ vanish extremely fast as $r \rightarrow r_c$. A more stringent test of the identity of the two transition points would be valuable.

We have also examined the observables discussed here in 1 + 1D circuits, motivated by the fact that, since they do not require us to introduce a spatial bipartition of the system, they avoid introducing a length scale that is smaller than the system size. We will report on these findings elsewhere.

F. Replicas and field theories

A key question is whether useful continuum field theories can be written for the MPT and FMPT, and also for entanglement transitions in (reasonably generic [70]) random tensor networks. This question has not been resolved,

despite progress on mapping the quantum problems to effective “classical” lattice models [13,14,34–36,42,43]. A basic issue is the need to handle disorder. The most familiar approach to this is to use the replica trick [13,14,34,36]. (In this section we use N to denote the number of replicas: this should not be confused with the number of physical spins in the previous sections.) However the complicated N dependence of the interactions makes it unclear *a priori* how to coarse grain these effective lattice models.

In Sec. VI we start by reviewing the approach of mapping circuits and tensor networks to effective lattice models for permutations. We discuss coarse graining of such models in a heuristic way. We then suggest an alternative way of thinking about effective statistical mechanics models for circuits (motivated by a physical picture for the emergence of permutations, in terms of phase cancellation in sums over Feynman histories [34,43]). This picture connects entanglement transitions to approaches familiar from disordered magnetism, the random-field Ising model, spin glasses, etc. [71].

With this motivation, we construct the simplest Lagrangians that capture the global symmetry associated with the replica formulation [34,36,43], which we denote

$$G_N \equiv (S_N \times S_N) \rtimes \mathbb{Z}_2, \quad (14)$$

and which pass some basic consistency tests.

The limiting number of replicas N is distinct for the case of (i) the MPT and (ii) both the FMPT and the RTN [13,14,36]. For the FMPT and RTN we need to take $N \rightarrow 0$, as in standard quenched disorder problems. For the MPT, realizations are weighted by the additional Born-rule factor, which increases the number of replicas: we need to take $N \rightarrow 1$ [13,14]. Previously, properties in the vicinity of a fine-tuned point have been used to motivate the suggestion that all of these problems may have similar universal properties, despite the differing numbers of replicas [14]. However, we find that the simplest field theory candidates (which may of course be too simple) are strikingly different in the two different cases.

The Lagrangians we propose have the schematic forms

$$\begin{aligned} \mathcal{L}_X &= \sum_{ab} [(\partial X_{ab})^2 + \mu X_{ab}^2 + X_{ab}^3], \\ \mathcal{L}_Y &= \sum_{ab} [(\partial Y_{ab})^2 + \nu Y_{ab} + Y_{ab}^3] + \sum_{abcd} Y_{ab} F_{ab,cd} Y_{cd}. \end{aligned} \quad (15)$$

F is the tensor $F_{ab,cd} = \delta_{bd} + \delta_{ac}$. In these equations we suppress all coupling constants except the crucial one that drives the transition, denoted μ or ν . Both space and time derivatives are grouped together in the derivative term: in the case of the circuit there will in general be a nonuniversal speed v appearing, so that the derivative terms have the

form $(\partial_t X)^2 + v^2(\nabla X)^2$. The plus sign means that there is an emergent Euclidean, rather than Lorentzian, space-time symmetry [6,9,15] (see endnote [72]).

In the Lagrangian \mathcal{L}_X , the field X_{ab} is a real $N \times N$ matrix satisfying $\sum_a X_{ab} = 0$ and $\sum_b X_{ab} = 0$. It may be thought of (modulo a constant shift) as a coarse-grained permutation matrix. This Lagrangian is appropriate for the replica limit $N \rightarrow 1$. It has upper critical dimension $D = 6$ (this is the space-time dimension in the case of the circuit). This is a candidate Lagrangian for describing the MPT.

At first we might assume that the same Lagrangian \mathcal{L}_X for the measurement transition could be continued to the distinct limit $N \rightarrow 0$ in order to describe the random tensor network and the forced-measurement transition. We argue in Sec. VI that this is not the case. Instead, the simplest candidate for the FMPT and RTN is the Lagrangian \mathcal{L}_Y . Here, the field Y is a real $N \times N$ matrix, with $N \rightarrow 0$, that does *not* satisfy any constraints on its row and column sums. The upper critical dimension for this theory is the unexpectedly large value $D = 10$. See Sec. VI for further discussion.

We caution that these theories are conjectures based on symmetry considerations and certain limited consistency checks. Further investigation is required to determine whether they are in fact sufficient to describe the problems of interest. It is possible that more elaborate continuum descriptions are required, either for a particular microscopic model or in general.

Indeed, the trees described in Sec. IID, which have exponential order parameter scaling close to the critical point, appear to be one case that is not captured by \mathcal{L}_Y . (Contrary to the naive guess that the high-dimensional limit of the field theory and the tree would show similar “mean-field” critical scaling.) We defer an examination of the reason for this to a future work.

In Sec. VI we also present some results that are independent of the speculative field theories above. In particular, we use effective domain wall pictures to obtain the scaling within the phases (mentioned above in Sec. IIB).

We also briefly discuss the use of Ising toy models for the properties of the second Rényi entropy in measurement dynamics, pointing out that the formalisms of Ref. [34] or Ref. [43] allow these to be justified in certain strongly entangled regimes, rather than being regarded simply as toy models as in previous work. However, quenched disorder must be taken into account in the resulting Ising model. Additionally, the Ising picture breaks down close to the critical point (or in the disentangling phase) and also at long times.

Finally we discuss variants of the MPT, FMPT, and RTN phase transitions. We point out that quite different scaling obtains for models of free fermions subjected to measurements, as a result of continuous rather than discrete replica symmetry.

III. MINIMAL CUT PROBLEM

A natural starting point for understanding the MPT is to map the quantum circuit to a classical graph on which one can study a classical “minimal cut” optimization problem [6]. In this mapping there is a phase transition at the point where the graph percolates.

We think of this classical min-cut problem as a toy model for the generic quantum transition. In the circuits we study, the cost of the minimal cut gives the exact value of the (somewhat unphysical) zeroth Rényi entropy [73], S_0 . It also gives exact results for the other Rényi entropies in the limit of large local Hilbert space dimension (e.g., a large value of each spin), with Haar-random gates. But in general, the minimal cut is only an upper bound on the entanglement entropies S_n with $n \geq 1$. (There can be no quantum information propagated from the initial to the final time if the associated classical graph is disconnected; in this regime, the “cost” of the minimal cut vanishes.) The true “quantum” transition in general occurs at a smaller value of r than the classical transition discussed in this section (and in general has distinct universal properties). Despite this, the classical problem conveys some useful lessons.

Viewed as a graph, the circuit is a bond percolation configuration, as described below. The frequency of projective measurements determines the fraction of broken bonds in this percolation configuration. The minimal cut is the minimal number of additional bonds that must be severed in order that two parts of the boundary of the circuit, A and \bar{A} , no longer have any percolating path between them. This minimal cut is a unifying heuristic [35,51,67,74–76] for the entanglement of various objects, depending on how we choose A and \bar{A} . If these are taken to be two complementary subsets of the legs of the circuit at the final time, then the minimal cut gives the entanglement S_0 of a subset A of the spins in the final state quantum state, assuming the initial state is a product state. Here we are more interested in a minimal cut separating the top boundary of the circuit from the bottom. That is, A contains all the circuits “legs” at the final time, and \bar{A} all those at the initial time. This “horizontal” minimal cut is a measure of information transmitted from the initial to the final time, equal to the operator entanglement S_0 for the nonunitary time evolution operator V (Sec. II).

In the percolating regime, this horizontal minimal cut must sever a number of bonds that is extensive in the number of spins N , so that $S_0 \simeq sN$. The coefficient s is a “surface tension” [51] for the minimal cut, which vanishes continuously at the percolation threshold. In $1 + 1D$ this transition is conformally invariant. Many of the critical exponents, such as the correlation length exponent ν , are standard percolation exponents, while others are less familiar, since the minimal cut is an additional optimization problem built on top of the percolation configuration [6,77].

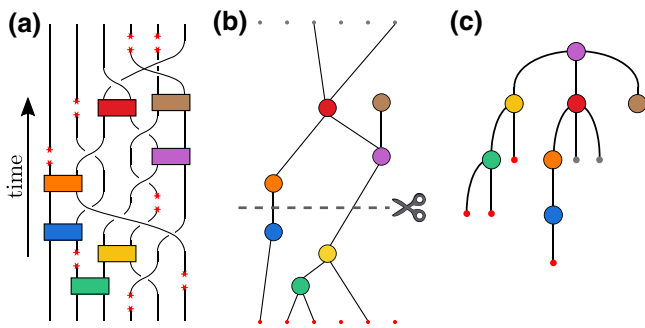


FIG. 4. (a) Example of a small unitary circuit with $N = 6$ spins. Black worldlines represent the evolution of a particular spin, with time proceeding vertically. Colored blocks indicate two-spin unitaries, and broken lines (marked with red crosses) indicate single-spin measurements. (b) The equivalent graph, with nodes representing unitaries [node of a particular color corresponds to the unitary of the same color in (a)] and edges representing unbroken segments of worldline. Small red (gray) circles denote the initial (final) time for a given spin. A possible minimal cut for this graph is shown by the dashed line: removing the two indicated edges disconnects the initial and final times. (c) The classical graph arranged as a tree, with the purple node used as a seed and generation number k proceeding downward. (This illustrative circuit forms a tree; in general the structure of a large circuit is only locally treelike.)

In the circuit without fixed spatial structure, where any qubit can couple randomly to any other, the location of the critical point, and the basic critical exponents, can be determined exactly, as we show in this section. These exponents also apply to the finite-dimensional minimal cut problem when the spatial dimensionality d is greater than or equal to 5 (Sec. III E). Interestingly, there is also reason to speculate that the exponents apply for some versions of the quantum measurement transition in high dimensions, even without the minimal cut approximation (see Sec. VI, where we discuss Landau theory for the measurement transition and entanglement transitions).

For all-to-all circuits, the classical percolation problem is defined as follows. The circuit defines a random graph, in which the nodes (vertices) correspond to unitaries and the edges are the sections of spin worldline that are not broken by measurements. In other words, an edge connects two nodes whenever (i) the two nodes correspond to successive unitaries in the time evolution of a particular spin; and (ii) that spin is not measured during the time in between the two unitaries. Figure 4(a) shows an example circuit, and Fig. 4(b) shows the corresponding graph. Each node has at most four edges connected to it, corresponding to the four legs of each unitary in Fig. 4(a). The minimal cut in the figure indicates an operator entanglement $S_0 = 2$ for this small circuit.

We take the number of spins to be very large, $N \gg 1$, while by definition the degree (connectivity) of each node

is only of order 1. In this situation, standard considerations [78] imply that the local structure of the graph is *treelike* on both sides of the percolation transition. Above the percolation transition closed loops do exist, but their length is of order $\ln N$.

A. Local tree structure and percolation

To relate the classical graph to a tree, imagine starting at an arbitrarily chosen “seed” unitary in the bulk of the circuit (far from the initial and final time boundaries) and tracing out its cluster: finding the nodes connected to the seed by an edge, then those connected to the seed by a path of length 2 edges, etc. In this way the cluster containing the seed may be arranged in a tree, with the seed at the top and subsequent generations of connected nodes below: see Fig. 4(c). We denote the generation number by k , with $k = 0$ being the seed.

The probability p that a given one of a unitary’s four possible edges is *absent* is equal to the probability that (as we travel along that segment of worldline) the spin undergoes a measurement before it is involved in another unitary. This probability is given by

$$p = \frac{r}{2 - r}. \quad (16)$$

The small circuit shown in Fig. 4 contains no loops. In general the circuit can contain loops. However, a subcluster of any finite size is guaranteed to be free of loops in the limit $N \rightarrow \infty$ (since the probability that two unitaries in generation k both connect to the same unitary in generation $k + 1$ is of order $1/N$).

To understand the location of the critical point, note that the average branching number of the tree (the average number of descendants of a given node with $k > 0$) is $3 \times (1 - p)$. The percolation transition in the graph occurs when the branching number is 1, i.e., at $p_c = 2/3$ (as also noted in Ref. [12]), or

$$r_c = \frac{4}{5}. \quad (17)$$

(In this section only, r_c denotes the classical transition point, $r_c = r_c^{\text{cl}}$.) When r is greater than r_c , all trees are finite even in the limit $N, t \rightarrow \infty$, where the graph itself is infinite: starting at a seed node, the tree inevitably dies out after a finite number of generations. Therefore, at $r > r_c$ all unitaries are in finite clusters; this is the nonpercolating phase. When $r < r_c$, however, there is a nonzero probability f_∞ that a tree continues forever, or rather until it includes a number of nodes proportional to N . In the percolation problem, f_∞ is the order parameter—the probability that the unitary lies in the infinite cluster. The critical exponent β for this order parameter is 1, which is

the mean-field value for percolation. A simple recursive treatment (Appendix A 1) shows that, close to r_c ,

$$f_\infty \simeq \frac{50}{3}(r_c - r)^\beta, \quad \beta = 1. \quad (18)$$

Note that the window of r in which we can hope to observe critical scaling, corresponding to $0 < f_\infty \ll 1$, is rather narrow as a result of the large (nonuniversal) prefactor in Eq. (18).

B. Effective 1D continuum theory

We now show that near the critical point the basic scaling variables for the percolation and minimal cut problems are

$$\frac{t}{N^{1/5}}, \quad \delta r N^{2/5}, \quad (19)$$

where $\delta r = r - r_c$ and t is, say, the temporal duration of the evolution. For example, the characteristic timescale for a large system at its critical point scales as $N^{1/5}$. In Secs. III C, III D and Appendices A 2–A 4 we show how these variables appear in scaling forms for the minimal cut and other observables. The critical exponents in Eq. (19) have also been obtained in independent work [57], by an approach that is complementary to the one below (Sec. C).

This problem is similar to one of crossover scaling, in which a system that is effectively very high dimensional on short timescales crosses over to one that is one-dimensional on long scales. This analogy can be used to obtain the above exponents, as we discuss in Sec. III E. This approach also sheds light on the quantum problem (Sec. VI). Here, however, we solve the classical problem directly.

To simplify the discussion, let us consider a percolation problem with the same basic features as the circuit, but with a simpler connectivity rule inspired by the Erdős-Rényi random graph [78]. This simplification does not change the universality class, as we show numerically in Appendix 3. The random graph we consider has a layered structure, with one layer for each timestep. This graph may be contrasted with one studied in Ref. [26], which maps a measurement transition in a class of “instantaneous quantum polynomial time” circuits to the percolation transition in an Erdős-Rényi graph without a time dimension.

We discretize the time t in integer steps. At each t we have N nodes, labeled (i, t) with $i = 1, \dots, N$. We allow edges only between sites in adjacent t layers, each edge being present with probability $b'/2N$, independently of the others. This scaling with N ensures that the average degree of a site, b' , is $\mathcal{O}(1)$, as in the circuit. It is easy to see by thinking about the local tree structure that the phase transition is at $b'_c = 1$. As in the circuit, connectivity is local in time, but there is no notion of spatial structure with a layer at a fixed time.

Classical percolation can be mapped to the Q -state Potts model in the limit $Q \rightarrow 1$ [79–82]. For our problem, the fact that each site couples to *all* the sites on the adjacent layers means that the Potts partition function simplifies after a Hubbard-Stratonovich transformation with a field $\Phi(t)$ that depends only on time. This transformation is shown in detail in Appendix A 2. The field Φ may be taken to be a $Q \times Q$ traceless diagonal matrix, on which Potts symmetry acts by permuting the diagonal components.

It is possible to take the continuum limit in a controlled way, to give an effective one-dimensional field theory. Close to the critical point, such that $b' - 1 = \delta b' \ll 1$, the partition function for this field theory is

$$Z = \int \mathcal{D}\Phi \exp\left(-\int dt \mathcal{L}\right), \quad (20)$$

with

$$\mathcal{L} = \frac{1}{4} \text{tr}(\partial_t \Phi)^2 - \frac{\delta b'}{2} \text{tr} \Phi^2 - \frac{1}{6\sqrt{N}} \text{tr} \Phi^3. \quad (21)$$

Modulo the values of the order-1 constants, we expect the same field theory to apply to the percolation model arising from the circuit.

The factor of $1/\sqrt{N}$ in Eq. (21) allows a long timescale and nontrivial scaling forms to emerge at the critical point $\delta b' = 0$, despite the fact that the effective field theory is one-dimensional. One-dimensionality implies that for any fixed N , correlations decay exponentially at sufficiently large t , but the timescale diverges with N .

The critical exponents for the minimal cut problem in the all-to-all circuit follow from the observation that the change of variables

$$\tilde{t} = \frac{t}{N^{1/5}}, \quad u = \delta r N^{2/5}, \quad \tilde{\Phi} = \frac{\Phi}{N^{1/10}}, \quad (22)$$

eliminates N from the action. Scaling forms for correlation functions follow from this fact together with the corresponding scalings for operators. We discuss some examples in the following subsection and in Appendix A 4.

We may also obtain these exponents from a crossover scaling argument if we assume that they are the same as those in a system that *does* have spatial structure, but with a very high spatial dimensionality d . This crossover is described in Sec. III E.

In fact, the exponents in Eq. (19) apply for any $d > 5$ (in an appropriate regime of timescales) with logarithmic corrections in $d = 5$. This is because $d = 5$ gives a total space-time dimension of 6, which is the upper critical dimension for percolation. This fact allows an even simpler mnemonic for the above exponents. Suppose for a moment that we are considering a graph with a regular lattice in spatial dimension $d = 5$, with $N = L^d = L^5$,

where L is the system size. $d = 5$ is the lowest dimension in which mean-field exponents apply (up to logarithms). In this picture, the first scaling variable above is simply t/L , corresponding to the dynamical exponent $z = 1$ in the five-dimensional theory, and the second scaling variable is $u = \delta r L^{1/\nu}$, with the mean-field correlation length exponent $\nu = 1/2$. In $d > 5$ we must also consider the dangerous irrelevance of the interaction term in the field theory [83] (which means that the relevant timescale is no longer t/L), but this term can be treated using a standard coarse-graining argument (Sec. III E).

C. The percolation probability

Before describing the minimal cut itself (Sec. III D), we first consider an observable that is simpler to study both analytically and numerically—namely, the probability P_{perc} of percolation between initial and final times in the classical graph. The value of $1 - P_{\text{perc}}$ is equivalent to the probability that the operator entanglement is exactly zero, since nonpercolation of the classical graph implies that the initial and final times are causally disconnected.

P_{perc} has scaling dimension zero, i.e., it has no power-law prefactor in N , so it is useful for numerical tests of the scaling defined by Eq. (19). In Appendix A 4 we present numerical results for two observables with nontrivial scaling dimension: namely, the probability of two nodes on either the same or opposite time boundaries being connected to the same cluster. We show that these observables are also described by the scaling variables in Eq. (19).

In the Potts language, P_{perc} is expressed in terms of the free-energy cost of twisted boundary conditions [84]. (In the 1D field theory this free energy involves boundary magnetic fields that are parametrically large in N ; this is discussed in Appendix A 2.) We obtain the scaling form:

$$P_{\text{perc}} = F\left(\frac{t}{N^{1/5}}, N^{2/5} \delta r\right). \quad (23)$$

(Here t denotes the full temporal duration of the dynamics.) First consider the critical point $r = r_c$, for which

$$P_{\text{perc}} = F_{\text{crit}}\left(\frac{t}{N^{1/5}}\right). \quad (24)$$

In principle we should obtain a scaling collapse simply by plotting P_{perc} as a function of the scaling argument. Practically speaking, however, the characteristic timescale $N^{1/5}$ is modest for the values of N we can access numerically, and it appears to be necessary to include a subleading correction. This correction is of a type that is generically present for nonperiodic boundary conditions, and corresponds to replacing the scaling variable with $(t - c_0)/N^{1/5}$, for a nonuniversal $\mathcal{O}(1)$ constant c_0 .

Figure 5 (inset) shows raw data for the percolation probability P_{perc} of the classical graph (for the all-to-all circuit)

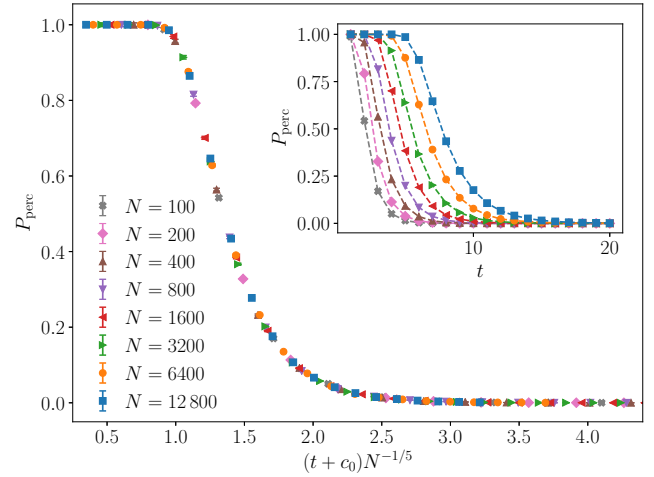


FIG. 5. The probability of percolation, P_{perc} , as a function of time for the classical circuit with the critical measurement rate $r = r_c = 4/5$. Different curves correspond to different system sizes N . The inset shows the raw data. In the main panel the time is rescaled by $N^{1/5}$ and a shift c_0 is introduced, with $c_0 \approx 1.3$. All data is averaged over 40 000 realizations.

as a function of time and N . As can be seen in the main panel, this data collapses onto a single curve when P_{perc} is plotted against $(t + c_0)/N^{1/5}$, where $c_0 \approx 1.3$.

At any fixed values of r and N , the probability P_{perc} decays exponentially with time t at large enough values of t . One can extract the associated decay time $\tau(r, N)$, which according to Eq. (23) has the scaling form

$$\tau(r, N) = N^{1/5} W(N^{2/5} \delta r). \quad (25)$$

This scaling is confirmed in Fig. 6. Figure 6 comprises a check of off-critical scaling close to r_c as well as the scaling at r_c that is shown in Fig. 5.

The decay time $\tau(r, N)$ of the percolation probability constitutes one way of defining a characteristic timescale over which information is able to propagate between the initial and final times. A key feature of the classical graph, which carries over to the quantum case, is that the timescale τ grows very rapidly with N within the entangling phase. At any fixed $r < r_c$, we can argue that as $N \rightarrow \infty$ the timescale τ grows as

$$\tau_{r < r_c} \sim \exp[a(r) \times N], \quad (26)$$

neglecting power-law prefactors. In the present classical problem, this exponential growth can be understood in terms of rare events that disconnect the cluster. Close to the transition we must have

$$a(r) \sim (\delta r)^{5/2} \quad (27)$$

in order to match the scaling form [85]. We expect that P_{perc} is close to 1 for $t \ll \tau$. This exponentially long

timescale can also be seen directly from the field theory in Eq. (21), in terms of “instantons” in the field theory (domain walls in time); see Appendix A 2.

As mentioned in the previous subsection, one can model the minimal cut problem using the simpler setup of a sequence of layered Erdős-Rényi random graphs. Within this model one can calculate the percolation probability P_{perc} and characteristic decay time τ . In Appendix 3 we show that this layered Erdős-Rényi model gives the same scaling behavior as in Figs. 5 and 6.

Percolation two-point functions give further information on the connectivity of the circuit. These are analyzed in Appendix A 4.

D. Scaling for the minimal cut

Because of the lack of spatial structure in the all-to-all model, it is natural to focus on the transmission of information between the initial and final times. One measure of this transmission is the operator entanglement of the linear, but nonunitary, operator V that defines the time evolution for a particular sequence of measurement outcomes (Sec. II).

In the minimal cut picture, the operator entanglement between initial and final times is the cost of the minimal cut through the circuit that separates the initial and final times [as illustrated in Fig. 4(b)]. We refer to this cost as S_0 (the Hartley entropy), although in some cases (including the limit of infinite local Hilbert space dimension, mentioned above) it is equal to the other Rényi entropies as well.

The behavior of S_0 is most interesting within the entangled phase, so let us consider some fixed $r < r_c$. As illustrated in the previous subsection, in this phase there is an

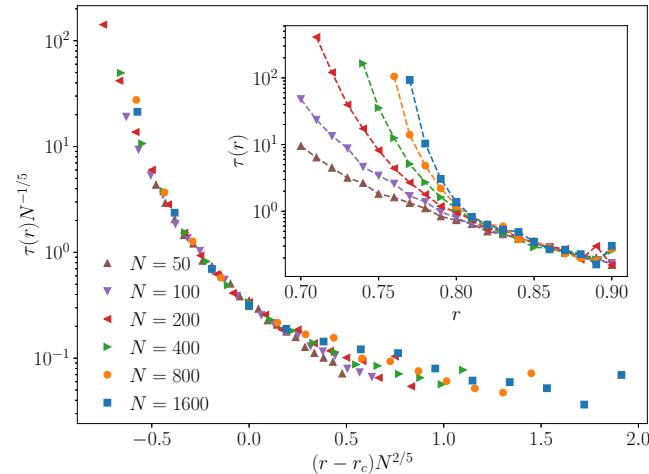


FIG. 6. The characteristic decay time τ for the percolation probability. The inset shows τ as a function of the measurement rate r for different values of the system size N . The main figure shows this same data plotted as a function of the scaling variables in Eq. (19).

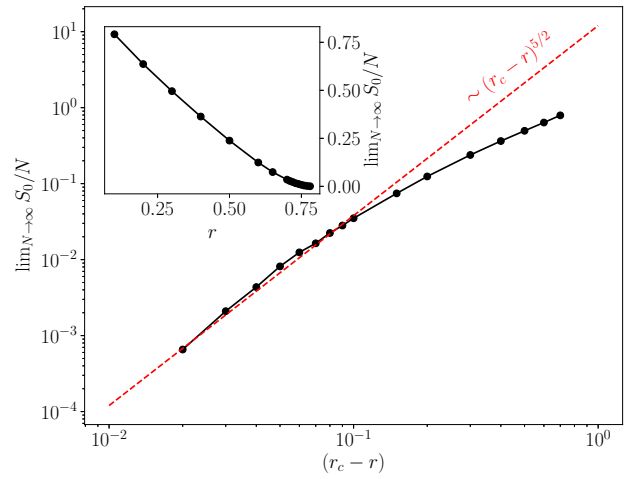


FIG. 7. The cost of the minimal cut, S_0 , divided by the system size N , extrapolated to the limit of $N \rightarrow \infty$. The main figure shows S_0/N as a function of r , while the inset shows S_0/N as a function of $r_c - r$ in double-logarithmic scale. The dashed red line shows the dependence $S_0/N \propto (r_c - r)^{5/2}$. Details of the extrapolation procedure are discussed in the text. Error bars are smaller than the symbol size.

exponentially large (in N) timescale over which the percolation probability is close to 1. Correspondingly, there is a parametrically large time range, corresponding to [86] $1 \ll \ln t \ll N$, over which S_0/N is approximately constant. The crudest picture for the subleading corrections gives (see endnote [87])

$$S_0/N = s - \mathcal{O}\left(\sqrt{N^{-1} \ln t}\right). \quad (28)$$

We refer to the range of times where subleading terms are negligible as the “plateau” in the entanglement. Over this large time window, the horizontal minimal cut has a well-defined cost per spin, s . This cost per spin s is the infinite-dimensional version of the line tension for the minimal cut in the 1 + 1D case or the surface tension in the 2 + 1D case [6]. These quantities all vanish at the critical point. In the plateau regime, the information *per spin* transmitted by the circuit is nonzero, up to an exponentially large time.

The inset of Fig. 7 shows this cost per spin in the entangled phase, as measured by a numerical simulation using the Ford-Fulkerson method [88]. The details of the extrapolation to large N are described below and in Appendix A 5.

Let us relate this minimal cut to the scaling theory close to the critical point. We expect the scaling form

$$S_0 = H\left(\frac{t}{N^{1/5}}, N^{2/5} \delta r\right). \quad (29)$$

(see Sec. III B). If we assume that within the entangled phase there is a time regime during which S_0 is extensive

in N and independent of time (i.e., independent of the first scaling variable above), then we obtain in this regime

$$S_0 = s(r)N, \quad (30)$$

with the entropy per spin $s(r)$ scaling as

$$s(r) \sim (r_c - r)^{5/2} \quad (r \lesssim r_c). \quad (31)$$

The main panel of Fig. 7 shows $s(r)$ close to the critical point on a double logarithmic scale. Though we cannot extract a clear power law from the data, it seems roughly consistent with the prediction, Eq. (31).

In order to numerically obtain the value of $s(r)$ for the plots above, we measure S_0 as a function of the time t and the system size N from simulations. For a fixed t , we find that $S_0(t, N)/N$ has a linear dependence on $1/\sqrt{N}$ at large N [in line with the simple picture in Eq. (28)]. This dependence allows us to estimate a value of $S_0(r)/N$ in the limit of $N \rightarrow \infty$ by extrapolating the linear relationship to $1/\sqrt{N} = 0$. Further details of this extrapolation procedure are presented in Appendix A 5.

E. Finite dimensions with $d \geq 5$

The scaling exponents identified in Secs. III B and III C also apply to the classical problem in a system with a regular spatial lattice (and unitaries applied only between nearest neighbors) in a large enough number of spatial dimensions d , as we now discuss. The total space-time dimension, $d + 1$, should be greater than 6, which is the upper critical dimension for percolation (in $d = 5$ we have the same exponents with additional logarithms).

We start with the standard Potts representation of percolation [79–82] in $d + 1$ dimensions. Suppressing all $\mathcal{O}(1)$ constants, as well as a nonuniversal velocity scale, a continuum action is

$$S = \int dt d^d x \operatorname{tr}[(\partial_t \phi)^2 + (\nabla \phi)^2 + \delta r \phi^2 + \phi^3]. \quad (32)$$

Here ϕ is a traceless diagonal $Q \times Q$ matrix, as in Sec. III B. Our system is of extent L in each of the spatial dimensions, with

$$N = L^d, \quad (33)$$

and extent $t \gg L$ in the time direction. We take the UV cutoff (“lattice spacing”) to be 1.

We coarse grain the system by a factor of order L , so that the spatial system size becomes comparable with the UV cutoff, and we have an effective 1D theory as far as correlations on scales $\gg L$ are concerned. Since the cubic coupling is irrelevant, with renormalization group (RG)

eigenvalue $y_3 = -(d - 5)/2$, it decreases during the flow, leading to (again we suppress order-1 constants):

$$S_{\text{eff}} = \int ds \operatorname{tr}[(\partial_s \phi')^2 + \delta r L^2 \phi'^2 + L^{-(d-5)/2} \phi'^3]. \quad (34)$$

Here s , the coarse-grained time coordinate, is equal to t/L , and $\phi' \sim L^{(d-1)/2} \phi$ from the scaling dimension of the field in $d + 1$ dimensions. If we now write the action in terms of t and $\Phi \equiv L^{1/2} \phi'$, we recover the form of the action in Eq. (21) with $N = L^d$.

Because of the dangerous irrelevance of ϕ^3 [83], a finite-dimensional model with $d > 5$ has two distinct large timescales,

$$L = N^{1/d} \quad \text{and} \quad L^{d/5} = N^{1/5}. \quad (35)$$

The shorter timescale (which is compressed to order 1 in the all-to-all model) marks the crossover between $(d + 1)$ -dimensional and one-dimensional scaling for correlation functions. The longer timescale is the one of more interest to us, and indicates the time at which the percolation probability starts to vary away from unity. This longer time becomes the characteristic critical timescale in the all-to-all model.

The scaling forms discussed above carry over to the present case ($5 < d < \infty$) with $N \rightarrow L^d$.

F. Lessons for the full quantum problem

So far we have discussed the (classical) minimal cut problem in all-to-all and high-dimensional circuits. *A priori*, one can expect the universal properties of the generic measurement transition to be different from those for the minimal cut transition: the minimal cut is only an exact representation of the entanglement in certain special cases (as described at the beginning of Sec. III). Nevertheless, as in 1 + 1D, the solution of the minimal cut problem provides more general lessons.

First, there are qualitative features that carry over to the generic problem. The most basic feature is the existence of a transition between a phase in which the operator entanglement $S(t)$ —the information propagated from the initial to the final time—decays quickly with time, and a phase in which an extensive value of entanglement, $S(t) \sim sN$, persists over a time that grows *exponentially* with the number N of spins. In Secs. V and I we demonstrate that these features carry over to the operator entanglement (as measured by the von Neumann or Rényi entropies $S_{n \geq 1}$), and related observables, in spin-1/2 circuits with measurements or forced measurements. Another generic feature is that close to the critical point, the scaling of the exponential timescale is tied to that of the plateau entanglement: $\ln \tau(r) \sim Ns(r)$ (Sec. I).

The minimal cut model also illustrates a possible relationship between the all-to-all case and the case of a

high-dimensional regular lattice. In the classical problem, the exponents of the all-to-all model are those of finite but high dimensions, once we take account of the dangerous irrelevance of interactions in high dimensions, which leads to a critical timescale L^{const} that is parametrically larger than the linear system size L (a timescale $\tau \sim L$ is what one would naively expect from $z = 1$ scaling). In Sec. VI we discuss similar crossovers in field theories for generic quantum models. However, we caution that our results in Sec. IV suggest more complex possibilities in the all-to-all systems.

Finally, we saw that in the classical problem, the percolation order parameter and the value of r_c could be obtained exactly by studying a simpler problem on a tree. In the next section, we propose that exact results for the full quantum version of the FMPT can also be obtained by studying trees: not only their classical connectivity, as here, but their “quantum” connectivity as defined by entanglement measures for tree tensor networks.

IV. ENTANGLEMENT TRANSITIONS IN QUANTUM TREES

A. Motivation for studying quantum trees

Locally, the all-to-all circuit has the structure of a tree (Sec. A). Viewing the circuit as a graph whose nodes are unitaries and whose edges are segments of spin worldline, the size of the smallest loops diverges when $N \rightarrow \infty$. This is true for all values of the measurement or projection rate, including deep in the entangled phase. We propose that this allows some exact results for the phase transition in the circuit, in certain cases (the FMPT), by studying the entanglement transition in a tree tensor network. As a by-product, we give exact results for general tree tensor networks.

Figure 8(left) is a schematic of the first $k = 3$ generations of the tree that is connected to one end of a link somewhere in the bulk of the circuit. For later convenience we use a slightly different convention for drawing the tree to that in Sec. A. Previously we “pruned off” all the branches below a projection operator, while in Fig. 8(left) we leave them in place, so that the number of descendants after k generations (the number of links at the base of the tree) is always 3^k . Each four-coordinated node in this figure, such as the one denoted t , includes a unitary, together possibly with projectors on its legs—we describe this below.

This tree is a tensor network. It has one free tensor index at the top and 3^k free indices at the bottom, and tensors t in the interior (built from a unitary and projectors). A basic way to characterize such a tensor network is via the amount of quantum information shared between apex and base. We can quantify this by the entanglement entropy between apex and base (Sec. C below). This language suggests analogous, but distinct, criteria for the classical and quantum transitions.

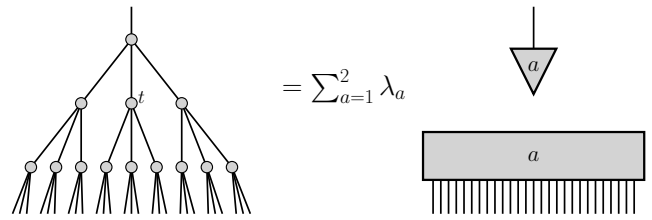


FIG. 8. A tree tensor network with $k = 3$ generations of nodes, and its singular value decomposition between apex and base. The individual node tensors are denoted by t [see, for example, Eq. (36)].

The *classical* percolation transition at r_c^{cl} (Sec. III above) has a simple interpretation in terms of the tree tensor network. For an asymptotically large tree, r_c^{cl} is the projection rate beyond which the apex and the base are guaranteed to be strictly disconnected by projectors. That is, once we go beyond the classical transition, the quantum information shared between apex and base vanishes for simple geometric reasons [89].

This suggests that we can also diagnose the *quantum* transition in the circuit, occurring at a value r_c^{qm} (we see below that $r_c^{\text{qm}} < r_c^{\text{cl}}$) using the properties of the tree. We show that the tree has a transition at a critical value r_c , which we conjecture is also the location of the critical point for the circuit ($r_c = r_c^{\text{qm}}$). For $r > r_c$ the amount of information shared between apex and base decreases exponentially with the number of tree generations, even though the apex and the base may not be disconnected in the trivial geometrical sense. For $r < r_c$, the von Neumann entanglement entropy between apex and base instead remains positive: $\lim_{k \rightarrow \infty} \langle S_1 \rangle > 0$.

Motivated by this connection between the circuit and trees, in this section we derive some universal results for entanglement transitions of tree tensor networks. We argue that the tree structure allows us to find the exact location of the critical point for the simplest version of the all-to-all circuit model exactly. In the language of Sec. II, this is the FMPT rather than the MPT. We explain in Sec. B immediately below why it is necessary for us to restrict to the FMPT in this section.

Tree tensor networks are also interesting quite apart from the connection to the all-to-all circuit [59–68]. They are instructive toy models for 1D wave functions with a scale-invariant entanglement structure [66,67], and they also allow efficient numerical tensor contraction algorithms [59]. Many of the results of the following subsections apply to more general disordered tree tensor networks that are unrelated to the circuit (see the discussion in Sec. I).

We obtain specific universal results for a broad class of trees that includes those arising in the FMPT circuit. These trees have bond dimension 2, and the probability distribution of the local tensors has a simple invariance property.

We also discuss, speculatively, what happens for trees with more general disorder distributions. Our conjectured continuum theory allows, *a priori*, for the the entanglement transition to be in distinct universality classes—a phenomenon analogous to a line of fixed points (there is an overview in Sec. D). Strikingly, for the class of trees that we study here, the transition is constrained to lie on a specific point on this line. It remains to be seen whether other points on the line can be obtained by varying the model.

Heuristically, these different possibilities for the tree transition can be related to different possibilities for the disentangled phase close to the transition. In the disentangled phase the entanglement between apex and base is exponentially small in k . But we can distinguish, in principle, between a “strong disorder” regime where this small amount of entanglement is (loosely speaking) dominated by a single path from apex to base, and a “weak disorder regime” where exponentially many paths through the tree contribute. For the tree tensor networks we study here, we show that the former (strong disorder) case applies. The possibility of these two regimes is due to the existence of a glass transition in the classical problem of a directed polymer on a tree [58]. We rely heavily on the methods developed in Ref. [58] for the directed polymer problem, which relate a linear recursion relation for the polymer’s partition function to a traveling wave equation.

B. Structure of tree tensor network

1. Generalities

The trees we consider have branching number three and bond dimension 2 for each bond (these are not essential restrictions). The four-index tensor t_{bcd}^a at a given node has bond index $a = 1, 2$ for the upper bond and b, c, d for the lower bonds. Below we describe the structure of t for the circuits we consider. We note that they fall within a special class of tree tensor networks with a simplifying feature, for which we be able to make strong statements.

Let us first consider trees such as Fig. 8(left) in general terms, without assuming that they arise from a circuit problem.

First, our analytical treatment will assume that the individual random tensors t_{bcd}^a for the nodes are statistically uncorrelated. This is important as it allows a simple recursive equation for the entanglement between top and bottom. We also take them to be identically distributed.

Second, for most of this section we assume that the probability distribution of the local tensors t_{bcd}^a has a simple invariance property. Namely, the distribution is invariant under multiplying an arbitrary $U(2)$ matrix u on any index: for example under

$$t_{bcd}^a \rightarrow t_{bcd}^a u d' d. \tag{36}$$

This feature simplifies the recursive equation: it means we can write a recursion for singular values alone, without having to keep track of singular vectors. In fact we need only a weaker condition: below the invariance property will hold for only the lower indices of t_{bcd}^a , which is sufficient.

The two assumptions above will be satisfied naturally for the circuit ensembles we consider, for example those built from Haar-random two-site gates. They turn out to lead to surprisingly strong constraints on the structure of the recursion.

Towards the end of our discussion of trees (Sec. I) we speculate about what happens when we relax the second condition.

2. Application to FMPT in circuit

In applying our results on trees to the transition in a circuit, the first assumption above (on the statistical independence of the node tensors) restricts us to the FMPT rather than the MPT.

Recall that for the MPT the measurement outcomes in the circuit are determined by Born’s rule. That means they have nontrivial statistics that depend on the random unitaries, violating the first assumption above. But in the FMPT the local projection operators are fixed *independently* of the choice of unitaries, not with Born’s rule. This means all the nodes of the tree are statistically independent, allowing a recursive statistical treatment. The nodes are described explicitly below. For the ensembles of two-site unitaries we study, it does not in fact matter *how* the directions of the local projections are fixed, so long as this is done independently of the realization of unitaries. For definiteness, we take all the local projectors to be onto the spin-up state.

To complete the specification of the circuit model, we just need to fix the distribution from which each two-site unitary U is drawn. (As in Sec. III, the rate at which projection operators is applied is r .)

3. Choice of ensemble of unitaries

The simplest choice is to take each U independently Haar-random in $U(4)$, i.e., drawn from the circular unitary ensemble.

$$\text{Ensemble 1: } U \sim \text{Haar}. \tag{37}$$

For this ensemble we find a “quantum” transition point $r_c = 0.749$ that is quite close to the classical one, $r_c^{\text{cl}} = 0.8$. In order to increase the separation between these transitions, we also consider a second ensemble with more weakly entangling gates. Unitaries in this second ensemble, referred to below as the “ Δt ensemble,” are of the

form

$$\text{Ensemble 2: } U = (V_1 \otimes V_2)U_{\text{fixed}}(W_1 \otimes W_2), \quad (38)$$

where V_1 , V_2 , W_1 , and W_2 are Haar-random one-site unitaries (ensuring the invariance property mentioned above), and U_{fixed} is a nonrandom, *fixed* unitary:

$$U_{\text{fixed}} = \exp(-i\Delta t H). \quad (39)$$

The parameter Δt controls the strength of the unitary, and consequently the position of the quantum transition (see Fig. 11). However, for concreteness we mostly refer to the fixed value $\Delta t = 0.3$. For H we use [6]

$$H = 0.3X_1X_2 + 0.2(X_1 + X_2) + 0.4Z_1Z_2 + 0.5(Z_1 + Z_2), \quad (40)$$

where X_i , Y_i , Z_i are the Pauli matrices on site i . These parameter values are not fine tuned, and we do not expect results to depend qualitatively on the precise values.

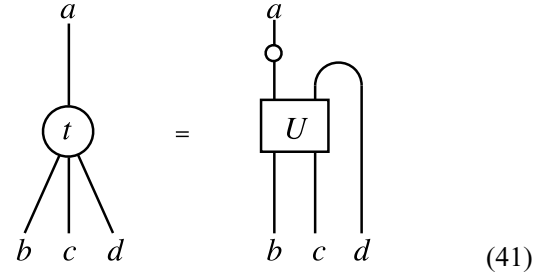
The recursive treatment for Z_k below applies to a more general family of distributions for the two-site unitaries for which the assumptions in Sec. 1 are obeyed. First, the distribution of U should be invariant under left or right multiplication by *single*-site unitaries [90]. Second, it should have the property of being statistically invariant under exchange of the two spins acted on by the unitary, and under transposition of the unitary. (This is less crucial, but simplifies the recursion [91]).

4. Node tensor in tree

Recall from Sec. III that we can “grow” the tree by starting at some seed location in the circuit and following links (segments of spin worldline) to form a cluster of unitaries at greater and greater distance from the seed. In fact, if we start on a link, we can think of it as a seed for *two* trees, one attached to each end of the link. It suffices to consider the properties of one of these trees separately. Truncating the tree at k generations gives a tensor network with a single bond at its apex and 3^k bonds at the base (we follow the convention in Sec. A where all branches are kept, even if they contain projections).

First consider a tree with no projections, where each node is a unitary. Now, when we include projections, each link of the tree has a probability p to contain a projection [92]. If a projection is present, we choose to incorporate it

into the node below the link. The node tensor is therefore



$$(41)$$

or in components (we write the row index of U as a superscript, and the column index as a subscript; both are multi-indices, since the unitary acts on two spins):

$$t_{bcd}^a = Q_d^a U_{bc}^{a'd}. \quad (42)$$

The matrix Q , shown as a circle in the picture, is either the identity, or the projector onto up, with probabilities $1 - p$ and p for each of the options. Recall that, in terms of the measurement rate [Eq. (16)],

$$p = \frac{r}{2 - r}. \quad (43)$$

In the case where the projector is present, we could simply prune off all the branches below it, but it is simpler to treat the geometry of the tree as fixed. Note that the distribution of Eq. (42) is invariant under multiplication of $U(2)$ matrices on any of the lower indices, as required in Sec. 1.

C. Entanglement between apex and base

We characterize the phase that the tree is in by the amount of quantum information shared between its apex and its base. Depending on the phase, this can either be exponentially small in the number k of generations of the tree, or it can be order 1 even for asymptotically large trees.

We can always think of such a tree tensor network as a wave function for a single spin at the apex and multiple spins at the base. The information shared between apex and base is then quantified by the entanglement entropy between top and base, or more formally, by the singular value decomposition when we partition the tensor network between the top and the base: see Fig. 8(right) [93].

Since the bond at the apex has a bond dimension of 2, there are only two singular values. After normalizing the tree, their squares sum to one, so we are in fact characterizing the tree by just a *single* number. We take this to be the square of the smaller singular value, and will denote it by Z_k for a tree with k generations of nodes:

$$Z_k = \lambda_{\min}^2. \quad (44)$$

The entanglement mentioned above is quantified by the Rényi entropies,

$$S_n = \frac{1}{1-n} \ln[Z_k^n + (1-Z_k)^n], \quad (45)$$

which at small Z_k are approximately

$$S_{n>1} \simeq \frac{n}{n-1} Z_k, \quad S_1 \simeq Z_k \left(\ln \frac{1}{Z_k} + 1 \right). \quad (46)$$

In the random tensor network Z_k is of course random. Its distribution can be obtained recursively, using the fact that a larger tree can be built up by combining subtrees.

D. Overview: classes of quantum tree

Let us summarize our basic conclusions for Z_k before getting into calculations. Depending on the location in the phase diagram, the random variable Z_k may have a broad distribution, and it will be vital to define its typical value using the average of $\ln Z_k$:

$$\ln Z_k^{\text{typ}} = \langle \ln Z_k \rangle_{\neq 0}. \quad (47)$$

We must condition on Z_k not being strictly zero in order to define the typical value [94].

For the quantum circuit with two-site unitaries and projections, the tree undergoes an entanglement transition at a critical value r_c . The value of r_c depends on the ensemble of unitaries, but a basic point is that it is strictly below the classical transition point for any ensemble satisfying our assumptions:

$$r_c < r_c^{\text{cl}}. \quad (48)$$

We compute the value of r_c analytically for the Haar circuit:

$$r_c = \frac{212 + 75\pi}{362 + 75\pi}. \quad (49)$$

This critical point at $r_c \simeq 0.749004$ lies not that far from the classical transition at $r_c^{\text{cl}} = 0.8$. For the $\Delta t = 0.3$ ensemble the spacing is increased,

$$r_c = 0.621(3). \quad (50)$$

In the disentangling phase the information shared between apex and base tends to zero exponentially with the size of the tree:

$$Z_k^{\text{typ}} \sim \exp(-|c_r|k) \quad \text{for } r > r_c, \quad (51)$$

with the ‘‘speed’’ $|c_r|$ vanishing linearly as $r \rightarrow r_c$.

In the entangling phase Z_k^{typ} is instead nonzero as $k \rightarrow \infty$, so that information is shared between apex and

base even in the limit of an infinitely large tree. This information becomes small as we approach the transition from the entangled side. The scaling is very rapid:

$$Z_\infty^{\text{typ}} \sim \exp\left(-\frac{C}{\sqrt{r_c - r}}\right) \quad \text{for } r \lesssim r_c, \quad (52)$$

with C a nonuniversal constant. The distribution of Z is also very broad when $r_c - r$ is small. For example $\langle Z_\infty \rangle \sim \sqrt{Z_\infty^{\text{typ}}}$, so that the mean is parametrically larger than the typical.

If we are right at the critical point, the value of Z decays more slowly with k than in the disentangled phase. A somewhat heuristic argument in Sec. 4 suggests

$$\ln Z_k^{\text{typ}} \sim -k^{1/3}. \quad (53)$$

The results above rely on an exact treatment of the linearized form of the recursion relation for Z_k , together with the conjecture that the effect of nonlinearity is captured by a simplified, analytically tractable model. Making this assumption (for which we provide numerical evidence), the results above hold for the trees derived from any forced-measurement circuit ensemble with the structure described in Sec. 3 (recall that we assumed invariance of the distribution of U under rotations on each leg).

In fact, they apply for the entanglement transition in *any* tree tensor network that obeys the two assumptions described in Sec. 1, in particular the $U(2)$ invariance property of the node tensor. In this context r is no longer interpreted as a measurement rate: instead it is any parameter characterizing t_{bcd}^a that can be used to drive the entanglement transition. However for the purposes of the discussion we use the notation appropriate to the forced-measurement circuit.

An analysis of random tree tensor networks outside the above class is a task for the future (see also Ref. [68]). However, our conjectured effective description suggests the interesting *possibility* that there may be multiple universality classes for the tree entanglement transition. The effective description includes a parameter $\Delta > 0$, which controls the scaling of Z_∞^{typ} near $r = r_c$. At first glance Δ is a nonuniversal parameter that will depend on the model. But, surprisingly, the $U(2)$ invariance property fixes $\Delta = 1/4$ at the entanglement transition. Equation (52) applies for $\Delta < 1$ in the effective description, but for $\Delta > 1$ there is instead power-law scaling of the ‘‘order parameter’’ Z close to the transition. For completeness, we solve the effective model in this $\Delta > 1$ regime also. We find a regime with a variable exponent, and a regime where this exponent is pinned to 1:

$$Z_\infty^{\text{typ}} \sim (r_c - r)^{\frac{1}{\Delta-1}} \quad \text{for } 1 < \Delta < 2. \quad (54)$$

$$Z_\infty^{\text{typ}} \sim (r_c - r), \quad \text{for } \Delta > 2. \quad (55)$$

However, at present it is unclear whether these regimes of the effective model can be accessed by any tensor network, or whether they exist only in the effective model (Sec. I).

E. Recursion relation for singular values

Let us think of the tree as a quantum state for a spin at the top and 3^k spins at the base: this is just to fix notation for bras and kets. We may write its Schmidt (singular value) decomposition:

$$T = \sum_{i=1,2} \lambda_i |i\rangle_{\text{top}} |i\rangle_{\text{bottom}}. \quad (56)$$

The states are Schmidt states in the appropriate Hilbert spaces (the second ket lives in the $2^{(3^k)}$ -dimensional Hilbert space associated with the base). In the problem we are studying, the overall normalization of the tree is not important, so we always take the Schmidt singular values to be normalized: $\lambda_1^2 + \lambda_2^2 = 1$.

Given three trees T_k , T'_k , and T''_k , each of k generations, we may form a tree T_{k+1} of $k+1$ generations by attaching T_k , T'_k , and T''_k to the base of the t node shown in Eq. (41). The statistical invariance of U under single-site rotations means that we are free to take the Schmidt states $|i\rangle_{\text{top}}$ [Eq. (56)] for T_k , T'_k , and T''_k to be simply the two basis states (up and down spin states), which we denote $|1\rangle_{\text{top}}$ and $|2\rangle_{\text{top}}$. Then

$$T_{k+1} = \sum_{a,b,c,d=1,2} t_{bcd}^a \lambda_b \lambda'_c \lambda''_d |a\rangle_{\text{top}} |bcd\rangle_{\text{bottom}}. \quad (57)$$

Here λ , λ' , and λ'' are singular values for T_k , T'_k , and T''_k , $\{|a\rangle_{\text{top}}\}$ are computational basis states, and $\{|bcd\rangle_{\text{bottom}}\}$ is a set of eight orthonormal states associated with the base of the full tree, formed from the Schmidt states of the three subtrees. Equivalently, in this basis,

$$(T_{k+1})_{bcd}^a = t_{bcd}^a \lambda_b \lambda'_c \lambda''_d. \quad (58)$$

It is straightforward to compute the normalized singular values of T_{k+1} [95]. Let us denote the smaller singular value squared of T_{k+1} by Z_{k+1} . If t includes the projector, then trivially

$$Z_{k+1} = 0. \quad (59)$$

We write the other case explicitly for completeness, though we need only a simple limit of it:

$$\begin{aligned} & Z_{k+1}^2 + (1 - Z_{k+1})^2 \\ &= \frac{\sum U_{bc}^{ad} (U_{bc}^{a'd'})^* U_{ef}^{a'g} (U_{ef}^{ag})^* \lambda_b^2 \lambda_c^2 \lambda_c'^2 \lambda_f'^2 \lambda_d'^2 \lambda_g'^2}{(\sum |U_{bc}^{ad}|^2 \lambda_b^2 \lambda_c^2 \lambda_d'^2)^2}. \end{aligned} \quad (60)$$

We are interested in a transition between a phase where Z_k vanishes as $k \rightarrow \infty$, and a phase where the typical value

of Z_k remains positive in this limit. Even in this phase, if we are close to the phase transition, this typical value of Z_k is small. Therefore, to understand the critical properties we can study the recursion relation in the regime where the minimal singular values are close to 0 for all the trees. We order the singular values of any tree such that $\lambda_2 \leq \lambda_1$, and define $Z = \lambda_{\min}^2 = \lambda_2^2$ for each tree.

The first step is to examine the linearized recursion relation. Taking Eqs. (59) and (60) to order Z ,

$$Z_{k+1} = \begin{cases} A_1 Z_k + A_2 Z'_k + A_3 Z''_k & \text{with probability } 1 - p \\ 0 & \text{with probability } p. \end{cases} \quad (61)$$

Here A_i are three positive constants that depend on the random unitary:

$$\begin{aligned} A_1 &= \frac{|U_{11}^{11} U_{21}^{21} - U_{21}^{11} U_{11}^{11}|^2}{(|U_{11}^{11}|^2 + |U_{21}^{21}|^2)^2}, \\ A_2 &= \frac{|U_{11}^{11} U_{12}^{21} - U_{12}^{11} U_{11}^{11}|^2}{(|U_{11}^{11}|^2 + |U_{21}^{21}|^2)^2}, \\ A_3 &= \frac{|U_{11}^{11} U_{11}^{22} - U_{11}^{12} U_{11}^{12}|^2}{(|U_{11}^{11}|^2 + |U_{21}^{21}|^2)^2}. \end{aligned} \quad (62)$$

Analogous formulas hold for more general choices of the node tensor t , see Sec. I.

Let us consider the meaning of this equation. Z_k , Z'_k , and Z''_k refer to trees of the same size, so they are drawn from the same probability distribution. The recursion relation then defines the probability distribution for a variable Z_{k+1} at the next level in the hierarchy. This defines a sequence of probability distributions $P_k(Z)$ for increasing k . The initial condition at the lowest level of the hierarchy is $Z_0 = 1/2$ (for a single bond, the two singular values are equal), i.e., $P_0(Z) = \delta(Z - 1/2)$. (This initial condition is far outside the linear regime, but close to the transition, Z_k becomes small at large k . The specific choice of the initial condition in the linear tree is unimportant as long as it is nonzero and positive.)

The linearized recursion relation, Eq. (61), is crucial. It is sufficient to obtain the exact location of the entanglement transition, although we need to add nonlinearity to understand what happens close to this transition in the entangled phase.

Let us collect here some properties of A_i that will be useful below. It turns out that the statistical invariance of U under single-site rotations allows some exact statements, regardless of the precise choice of distribution for U . We demonstrate these in Appendix 2. In particular, we need the following identities, which hold for all $i = 1, 2, 3$ (so

TABLE I. Some averages of the A_i that are used in the text.

| Quantity | Haar | $\Delta t = 0.3$ |
|---|-----------------------|------------------|
| $\langle A_i \rangle$ | 1 | 1 |
| $\langle \sqrt{A_i} \log A_i \rangle$ | 0 | 0 |
| $\langle \sqrt{A_1} \rangle$ | 64/75 | 0.981 |
| $\langle \sqrt{A_2} \rangle$ | 64/75 | 0.420 |
| $\langle \sqrt{A_3} \rangle$ | $\pi/4$ | 0.419 |
| $\langle \sqrt{A_1} (\log A_1)^2 \rangle$ | $\frac{17408}{16875}$ | 0.227 |
| $\langle \sqrt{A_2} (\log A_2)^2 \rangle$ | $\frac{17408}{16875}$ | 1.429 |
| $\langle \sqrt{A_3} (\log A_3)^2 \rangle$ | $\pi^3/4 - 2\pi$ | 1.428 |

long as U is nontrivially entangling with probability 1):

$$\langle A_i \rangle = 1, \quad \langle A_i^{1/2} \ln A_i \rangle = 0. \quad (63)$$

The first of these is at first sight surprising, since if the unitary U is the identity [for example, if $\Delta t \rightarrow 0$ for the distribution in Eqs. (38) and (39)] then A_2 and A_3 are exactly equal to zero [96]. However, this is a singular limit for $\langle A_i \rangle$, see below.

In the case where U is a Haar-random $U(4)$ matrix, we can obtain more general analytic results (Appendix 2):

$$\langle A_1^\lambda \rangle = \langle A_2^\lambda \rangle = \frac{12}{12 + 8\lambda - 7\lambda^2 - 2\lambda^3 + \lambda^4}, \quad (64)$$

$$\langle A_3^\lambda \rangle = \frac{\pi\lambda(1-\lambda)}{\sin(\pi\lambda)}. \quad (65)$$

To determine the location of the phase transition we need the special cases $\langle A_i^{1/2} \rangle$. These are given in Table I (Appendix 2) for both Haar and $\Delta t = 0.3$ ensembles.

The asymptotics of the probability distributions of A_i are also obtained in Appendix 2. These three variables are correlated, but here we discuss only the marginal distribution of a given one. Let us define

$$V_i \equiv \ln A_i. \quad (66)$$

For any generic distribution of U , the tails of the V_i distribution are exponential:

$$P(V_i) dV_i \sim \begin{cases} e^{-2V_i} dV_i & V_i \gg 0, \\ e^{-|V_i|} dV_i & V_i \ll 0. \end{cases} \quad (67)$$

If the distribution of unitaries is taken to be weakly entangling, for example if $\Delta t \ll 1$ in Eq. (39), then the right-hand tail of the distribution has an intermediate part, extending over the range $\ln \Delta t^2 \ll V_i \ll \ln \Delta t^{-2}$, that decays with the smaller exponent $-1/2$ (Appendix 2). This slowly decaying tail, cut off at a parametrically large V_i , is responsible for the failure of the limit $\Delta t \rightarrow 0$ to commute with the average in Eq. (63) that was mentioned above.

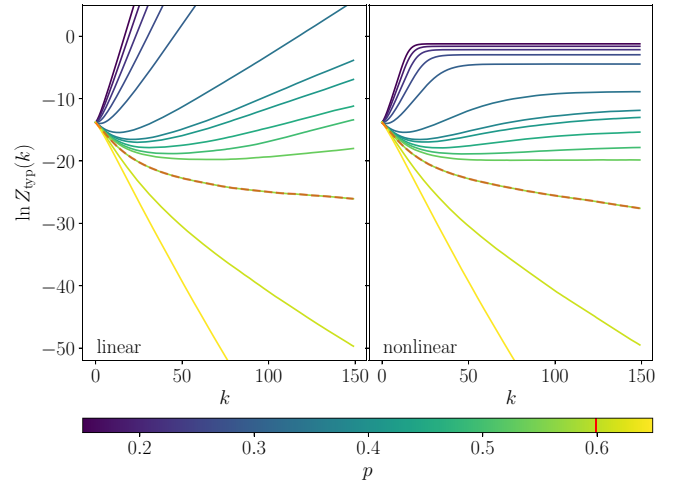


FIG. 9. Evolution of the typical value of Z with the depth k of the tree, for the linear recursion (left) and the nonlinear recursion (right), for various values of p . The evolution is similar in the two cases for $p > p_c$, but for $p < p_c$, the nonlinearity causes Z^{typ} to saturate. The initial value $Z_0 = 10^{-6}$ is used.

F. Linearized recursion relation

The most basic question about the linearized recursion, Eq. (61), is whether the typical value of Z is exponentially growing or exponentially shrinking at large k [58]. We may define an exponential growth speed c_p :

$$\ln Z_k^{\text{typ}} \sim c_p k. \quad (68)$$

In this section we usually use p as the parameter, rather than the equivalent r , Eq. (43), since p has a more direct interpretation in terms of the tree.

Define the point p_c to separate a regime of exponential growth, which we see occurs for $p < p_c$, from a regime of exponential decay at larger p :

$$c_{p_c} = 0 \quad (\text{definition of } p_c). \quad (69)$$

We see that p_c is precisely the location of the entanglement phase transition for the tree. When the linear recursion predicts that $Z^{\text{typ}} \rightarrow 0$ at large k , this remains true when higher powers of Z are included in the recursion. On the other hand, when the linear recursion predicts that $Z^{\text{typ}} \rightarrow \infty$ at large k , then the nonlinear terms in the recursion replace “ ∞ ” with a finite value, in a universal manner that we discuss in Sec. H.

This is illustrated using simulations for the case of Haar-random unitaries in Fig. 9. This compares Z_k^{typ} for the linear and nonlinear recursion relations. In the linear case, evolution follows $\ln Z_k^{\text{typ}} \propto k$ at large k , for all p . In the nonlinear case, this is only true for $p > p_c$. Details of these simulations are described in Appendix C 1.

The speed c_p can be extracted using the method of Ref. [58], which relates the linear recursion to a traveling wave

problem. Define the generating function

$$G_k(x) = \langle \exp(-e^{-x} Z_k) \rangle, \quad (70)$$

where the average is over Z_k . The recursion relation, Eq. (61), then becomes

$$G_{k+1}(x) = p + (1-p)\langle G_k(x - V_1) G_k(x - V_2) G_k(x - V_3) \rangle, \quad (71)$$

where the remaining average is only over the $V_i = \ln A_i$ defined in Eq. (66). [The fact that the V_i appear additively in the arguments of the generating functions here is the reason why the generating function in Eq. (70) is usually written with the double exponential.]

It may be helpful to think of $G_k(x)$, defined in Eq. (70), as a smeared version of the cumulative probability distribution for $\ln Z$. This definition shows that for $x \gg \ln Z^{\text{typ}}$, G_k plateaus at the value 1, while for $x \ll \ln Z^{\text{typ}}$, G_k plateaus at the probability of Z_k being exactly zero [97]. G_k has a ‘‘front’’ at

$$x_{\text{front}}(k) = \ln Z_k^{\text{typ}} + o(k) \quad (72)$$

that interpolates between these two plateaus. It is useful to think of x as a fictitious spatial coordinate, and of k as fictitious time coordinate [58]. Then, at late time, this front propagates as a traveling wave with speed c_p , and obeys the traveling-wave ansatz:

$$G_k(x) = G^{(\lambda)}[x - v_p(\lambda)k]. \quad (73)$$

The wave speed $v_p(\lambda)$ depends on a parameter λ of the solution $G^{(\lambda)}$. This parameter, which must be determined, is the exponential decay constant of $G^{(\lambda)}$ at large argument [58]:

$$G^{(\lambda)}(u) \sim 1 - \alpha e^{-\lambda u}. \quad (74)$$

Substituting this form into Eq. (71) gives an explicit formula for the speed $v(\lambda)$ of the traveling-wave solution with a given λ :

$$v_p(\lambda) = \frac{1}{\lambda} \ln \left[(1-p) (\langle A_1^\lambda \rangle + \langle A_2^\lambda \rangle + \langle A_3^\lambda \rangle) \right]. \quad (75)$$

We must then determine the correct value of λ , i.e., which traveling-wave solution the initial condition converges to. This is done by standard considerations of velocity selection for traveling waves [58,98].

In outline, there is a privileged *minimal speed* traveling wave defined by the parameter value $\lambda = \lambda_*$ where $v_p(\lambda)$ is minimal:

$$v_p^{\text{min}} \equiv v_p(\lambda_*), \quad v'_p(\lambda_*) = 0. \quad (76)$$

G_k will converge to this minimal speed solution if λ_* is less than 1 [99], while it will converge to the solution

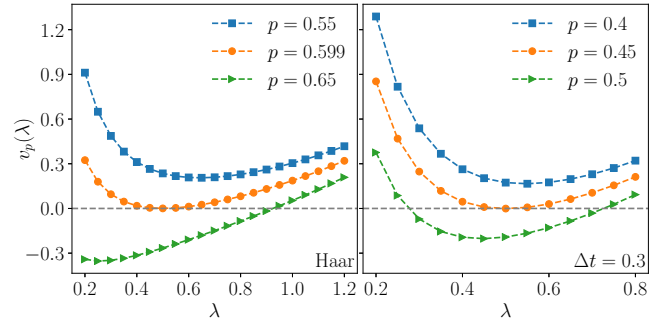


FIG. 10. Velocity $v_p(\lambda)$ vs λ for three values of p (above, and below the critical point). Left panel is for Haar evolution and right is for $\Delta t = 0.3$.

with $\lambda = 1$ if $\lambda_* > 1$. In the latter case the speed is $v_p(1)$, which we refer to as the ‘‘annealed’’ value of the speed (for reasons described in Sec. G):

$$v_p^{\text{ann}} \equiv v_p(1). \quad (77)$$

Therefore, the desired exponential growth rate is given for any p by

$$c_p = \begin{cases} v_p^{\text{min}} & \text{if } \lambda_* < 1, \\ v_p^{\text{ann}} & \text{if } \lambda_* > 1. \end{cases} \quad (78)$$

Recall that λ_* is determined using Eq. (75), via $v'_p(\lambda_*) = 0$, so it depends on p .

The above Eq. (78) can lead to a nonanalyticity in c_p as p is varied. This has a meaning in terms of the statistical mechanics of the linearized recursion relation [58], which we review in Sec. G. For now we simply note that, for the present class of circuits [100] the first line in Eq. (78) is always the one that applies for p close to p_c . This is shown in Sec. G. Given this, p_c is determined by solving

$$v_{p_c}(\lambda_*) = 0, \quad v'_{p_c}(\lambda_*) = 0, \quad (79)$$

for λ_* and p_c .

Figure 10 shows $v_p(\lambda)$, defined in Eq. (75), for the Haar tree and the $\Delta t = 0.3$ tree, in the vicinity of their respective p_c values. [Numerically, these are obtained by simple averages using a single tensor. In the Haar case, Eqs. (64) and (65) also give the exact form.] c_p is given by the minimal value of the curve, $c_p = v_p(\lambda_*)$, which passes through zero at $p = p_c$.

This can be used to determine p_c numerically, but in fact further analytical progress is possible. Using the definition of $v_p(\lambda)$ in Eq. (75), Eqs. (79) reduce to

$$\sum_{i=1}^3 \langle A_i^{\lambda_*} \rangle = \frac{1}{1-p_c}, \quad \sum_{i=1}^3 \langle A_i^{\lambda_*} \ln A_i \rangle = 0. \quad (80)$$

Remarkably, the second identity in Eq. (63) shows that the solution is always at $\lambda_* = 1/2$, for any ensemble of

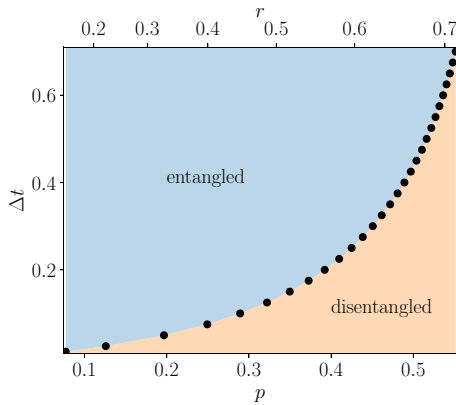


FIG. 11. Phase diagram of the quantum tree generated by the unitary gate in Eq. (39), in the space of Δt and r .

unitaries satisfying our assumptions. This fact gives an explicit expression for p_c as a simple average for the local node tensor,

$$p_c = 1 - \frac{1}{\sum_{i=1}^3 \langle A_i^{1/2} \rangle}. \quad (81)$$

This may be evaluated analytically for the Haar case [Eqs. (64) and (65)], giving $p_c = (212 + 75\pi)/(512 + 75\pi)$ (equivalent to the r_c value quoted in Sec. D) and numerically for the Δt ensemble. The location of the critical point in the Δt ensemble is shown for various values of Δt in Fig. 11.

G. Aside: glass transition in linear recursion

The canonical example of linear recursion relations like Eq. (61) is the problem of the directed polymer on a tree [58]: see Fig. 12. In the disentangled phase, where the linear treatment is valid at large k , this gives another interpretation of the singular-value-squared Z_k as a sum over paths through the tensor network. Here we briefly review this mapping and use it to clarify which of the regimes in Eq. (78) is relevant. This subsection is not essential to the subsequent development.

Within the linear approximation Eq. (61), Z_k is exactly equal to the partition function of a polymer that lies along a path from the top to the bottom of a tree of depth k , as in Fig. 12. We view $-V_1, -V_2, -V_3$ in Eq. (66) as random potentials on the three bonds below a given node. The energy of the polymer is the sum of the potentials for the bonds it visits:

$$Z_k = \sum_{\text{paths}} e^{\sum_{\text{on path}} V_{\text{bond}}}. \quad (82)$$

This is easily seen to satisfy the recursive Eq. (61). There are minor differences from the standard polymer model.

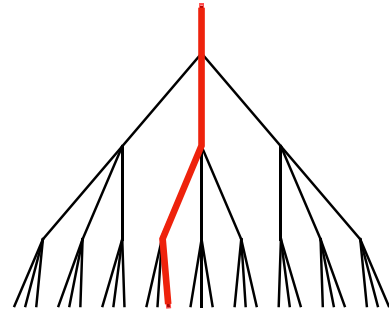


FIG. 12. Schematic of a directed polymer on a tree.

First, if $p > 0$, there are some bonds that the polymer cannot visit, where $A = 0$ or $V = -\infty$ (these bonds and the subtrees below them can simply be removed). Second, the V s have a nontrivial distribution, with links that share the same parent node having correlated potentials.

The polymer can be in either a glass phase or a paramagnetic phase [58]. These are distinct thermodynamic phases in the polymer problem, but to avoid confusion we refer to them as “regimes,” because they do not correspond to distinct phases of the entanglement problem. (The distinction between the glass and paramagnet is a feature of the linearized problem only, and is unrelated to the distinction between entangled and disentangled phases.)

The glass obtains when the pinning effect of disorder on the polymer defeats the depinning effect of entropy. Usually the glass would be entered by decreasing the temperature (increasing the scale of V). Here we increase the strength of disorder by increasing p . In the paramagnetic regime the polymer has extensive entropy (proportional to k) while in the glass the entropy per unit length vanishes.

The glass and paramagnet regimes have a simple translation to the language of the traveling wave (Sec. IV F), which we state only [58]. The polymer is in the glass regime if $\lambda_* < 1$, and in the paramagnetic regime if $\lambda_* > 1$ [58]. These correspond to the two lines in Eq. (78) for the growth rate c_p , which is simply (minus) the free energy per unit length of the polymer.

In our problem, the entanglement transition necessarily takes place in the glass regime of the linear recursion, essentially because of the fact that $\langle A_i \rangle = 1$. Let us give an intuitive picture.

To begin with, imagine that the polymer is in the paramagnetic regime. In this regime (but not in the glass [101]) the “annealed” expression for the free energy growth rate c_p applies [the second line of Eq. (78)]. This expression is in fact just the annealed approximation to the free energy, in which we average the partition function of the polymer, $\langle Z_k \rangle$, instead of averaging its logarithm. In the present linearized problem this gives

$$v_p^{\text{ann}} = \ln(3[1-p]), \quad (83)$$

using $\langle A_i \rangle = 1$.

Recall that the entanglement transition is at the value of p where $c_p = 0$. We see from Eq. (83) that if the polymer is in the paramagnetic regime in the vicinity of p_c , then the entanglement transition would coincide with the classical percolation transition at $p_c^{\text{classical}} = 2/3$!

We can see that this is inconsistent as follows. Consider the structure of large trees when we approach the classical percolation transition at $p_c^{\text{classical}} = 2/3$ from below. After deleting subtrees that terminate before reaching the base [102], a large tree with $Z \neq 0$ is made up of one-dimensional chains connected by branching events. Close to the classical transition, the typical length of one of these 1D chains grows like $(2/3 - p)^{-1}$ [103]. Treating them as renormalized bonds in the polymer problem, one may check that the effective disorder strength on these renormalized bonds grows without bound as they get longer. This increasing disorder strength implies that we must enter the glass regime before we get to the classical transition. That is, either the linear recursion relation is in the glass regime for all p , or it is in the glass regime for all $p > p_{\text{glass}}$ for some $p_{\text{glass}} < 2/3$.

When the polymer is in the glass phase, c_p is strictly smaller than the annealed approximation above [Eq. (78)]. Therefore, c_p in fact hits zero at a smaller value of p than v_p^{ann} does. In other words, p_c is strictly smaller than p_c^{cl} .

The value of p_{glass} is determined by the equation $v_p'(1) = 0$. For the $\Delta t = 0.3$ ensemble the value of p_{glass} is evaluated numerically and found to be negative, indicating that this ensemble is always in the glassy phase. For the Haar ensemble, $p_{\text{glass}} = (3 - e^{7/9})/3 \approx 0.274$ [from Eqs. (64), (65), and (75)]. But since this value lies inside the entangled phase, where the linearized recursion is not valid, we do not expect that the glass transition is physically significant for the tensor network.

The arguments here, showing that the entanglement transition must take place within the glass regime of the linear recursion, extend to the class of tree tensor networks described in Sec. 1. The possibility of other universality classes of entanglement phase transition for other kinds of quantum trees is discussed in Sec. 1.

H. Including the nonlinearity

Having understood the linear approximation to the recursion relation for the singular value squared, Eq. (61), we must now consider the effect of nonlinearity. The nonlinearity is necessary to make sense of the entangled phase, where Z_k is of order 1, rather than being exponentially large in k as the linear equation would predict. Our aim in this section is to determine the scaling of Z close to the transition, on the entangled side. Our basic conclusions have already been summarized in Sec. D.

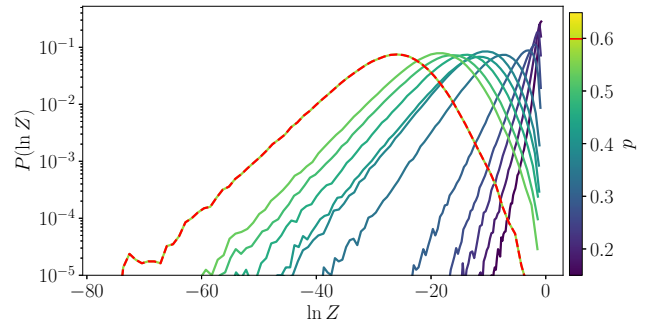


FIG. 13. Distribution of $\ln Z$ in a tree of $k = 150$ generations (for the full nonlinear problem). Here we remove instances where Z is exactly zero (which have a finite support due to the forced measurements).

I. Numerical results

Let us show numerical results before turning to an analytical treatment.

First, Fig. 13 shows the probability distribution of $\ln Z_k$ for the Haar ensemble (Sec. 3) in a tree of $k = 150$ generations, where we remove instances where Z is exactly zero [104]. Various values of p less than or equal to p_c are shown. The maximal possible value of Z_k is $1/2$: deep in the entangled phase the distribution is concentrated near this upper limit, but as we approach the critical point $\ln Z^{\text{typ}}$ moves to the left. The shape of the distribution also stabilizes. (In fact it approaches the shape for the linear problem, except on the right where Z is of order 1.)

Next, in Fig. 14 we show the scaling of Z^{typ} for both choices of the ensemble of unitaries (Sec. 3), close to the critical point. The analytic treatment below gives $\ln Z^{\text{typ}} \simeq -D/\sqrt{r_c - r}$, which corresponds to a straight line with slope $-1/2$ in the plot. This slope is indicated by the trend line. The data is consistent with this value of the exponent.

However, the value of the nonuniversal constant D that we extract from fitting this data is $D = 2.01$ for Haar and $D = 3.24$ for $\Delta t = 0.3$, which is far from that predicted below, for both ensembles. Experimenting with simpler toy models suggests that this may just be because of finite $r_c - r$ effects, i.e., not being close enough to r_c . The numerical method we use is afflicted by severe finite size effects (see Refs. [105–107] and Appendix C 1), associated with correctly sampling the right-hand tail of the distribution in Fig. 9, which mean we cannot approach too close to the critical point. Details of the numerical method are in Appendix C 1.

2. Nonlinear toy model

The nonlinear recursion relation in Eqs. (59) and (60) is not very approachable, even if expanded only to quadratic order. To make progress, we conjecture that the universal properties can be understood in a simpler model that

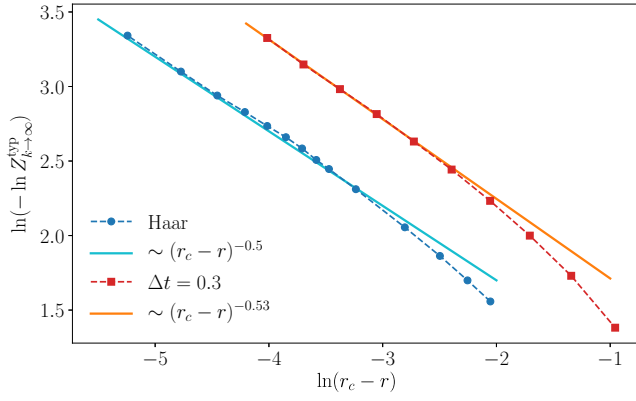


FIG. 14. Critical behavior for the Haar and $\Delta t = 0.3$ trees as obtained from Z_{600}^{typ} and Z_{700}^{typ} , for Haar and $\Delta t = 0.3$, respectively, for a pool size $N = 3 \times 10^4$. The data suggest $Z_{k \rightarrow \infty}^{\text{typ}} \sim e^{-D(r_c - r)^b}$ with the best fits for b being -0.5 and -0.53 for the Haar and $\Delta t = 0.3$ trees, respectively; the corresponding values of D are 2.01 and 3.24, respectively.

retains a few basic features. We study a recursion relation satisfying two requirements. First, it contains both linear terms and nonlinear terms of order Z^2 , which tend to suppress Z (naively, the terms of higher order than Z^2 should be negligible when we are parametrically close to the transition and $Z \ll 1$). Second, its linearized form is in the glass regime, as for the circuit. (Though in fact

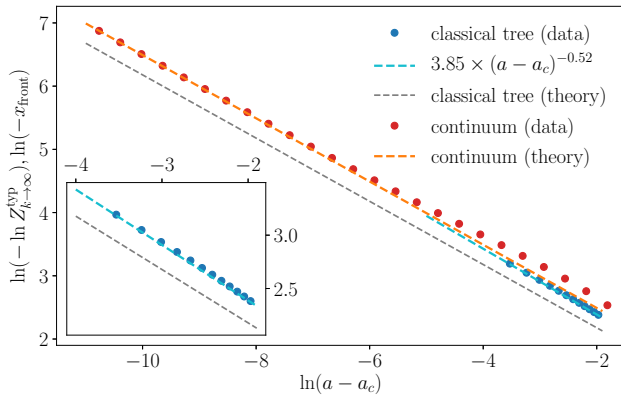


FIG. 15. Critical behavior of the classical nonlinear tree and the continuum equation in the glass phase. Similar to the quantum trees, the data for the classical tree (blue dots) suggests a scaling approximately $e^{-c(a_c - a)^b}$ for $Z_{k \rightarrow \infty}^{\text{typ}}$. The best fits are $b = -0.52$ and $c = 3.86$. For the parameters for the tree used, $p_0 = 0.2$, $p_1 = p_2 = 0.4$, $\epsilon = 1.411$, and $\gamma = 0.5$, the theoretical result shown in gray corresponds to $c = 3.247 \dots$ and $b = 0.5$. Data obtained by numerically solving the continuum equation (red dots) is in excellent agreement with the theoretical prediction close to the critical point, although some deviations are observed further away from the critical point. Inset: enlargement of the data for the classical tree.

we study both this case and the paramagnetic case for completeness.)

We first write down a toy model for a tree with a discrete generation number k , but as in Ref. [58] it will be convenient to take a continuum limit in k . We assume that this continuum limit preserves the universal properties, as is the case for the linear problem.

For the toy model, define the random variable Z_{k+1} at level $k+1$ in terms of a sum of ℓ random variables $Z_k^{(1)}, \dots, Z_k^{(\ell)}$ at level k . Here ℓ is the branching number of a node, and is taken to be random with a distribution p_ℓ for $\ell \geq 0$. The precise range allowed for ℓ is not important, so for simplicity we allow $\ell = 0, 1, 2$. We also include nonlinearity of strength γ , and a multiplicative random variable written as e^V , with Gaussian V :

$$Z_{k+1} = \left[e^V \sum_{i=1}^{\ell} Z_k^{(i)} \right] \exp \left\{ -\gamma \left[e^V \sum_{i=1}^{\ell} Z_k^{(i)} \right] \right\}. \quad (84)$$

This can be viewed as the composition of a linear transformation analogous to Eq. (61),

$$Z_{k+1} = e^V \sum_{i=1}^{\ell} Z_k^{(i)} \quad (85)$$

(but slightly simpler because we avoid having correlated random variables) and a nonlinear one,

$$Z_{k+1} \longrightarrow Z_{k+1} e^{-\gamma Z_{k+1}}. \quad (86)$$

The exponential form is arbitrary: for the continuum limit below it will anyway be sufficient to expand only to order γ , giving a quadratic recursion relation for Z . However, the above form guarantees that Z_{k+1} is positive for any input values, which is important for our numerical explorations.

We conjecture that by solving this simple nonlinear system we also capture universal scaling for the problem of interest (Sec. E).

3. Continuum traveling wave equation

The equation for the generating function [cf. Eq. (70)] that follows from expanding Eq. (84) to order γ is

$$G_{k+1}(x) = \exp(\gamma \partial_x e^x \partial_x) \sum_{\ell} p_{\ell} \langle G_k(x - V)^{\ell} \rangle_V, \quad (87)$$

where p_0 and p_2 are the probabilities of a termination and a branching, respectively. Now we take the continuum limit in the “time” k . When $\gamma = 0$, this gives the Fisher-KPP traveling wave equation [58]. We introduce a “time” step $\delta\tau$, which will be sent to zero and define $\tau = k\delta\tau$. The probabilities p_0 and p_2 are taken to be of order $\delta\tau$ (i.e., $p_1 = 1 - p_0 - p_2$ is close to 1) so that in the limit the tree

becomes a continuous time branching process. The parameter γ is taken of order $\delta\tau$ (note that the prefactor does not matter: it can be absorbed into the normalization of Z) and the strength of the random potential is also taken to vanish with $\delta\tau$. It is convenient to parameterize its mean and second moment as

$$\langle V \rangle = \frac{a}{2} \langle V^2 \rangle \quad \langle V^2 \rangle = 2b \times \delta\tau. \quad (88)$$

Finally we absorb some constants into the generating function by defining [108]:

$$G_k(x) = 1 - \frac{p_2 - p_0}{p_2} H_k(x). \quad (89)$$

The asymptotics of H may be taken to be

$$H(-\infty) = 1, \quad H(\infty) = 0. \quad (90)$$

After absorbing a constant into the definition of τ , and shifting x by a constant, H satisfies

$$\partial_\tau H = \partial_x [D(x)\partial_x - a]H + \Delta H(1 - H), \quad (91)$$

where the growth rate is $\Delta = 2(p_2 - p_0)/\langle V^2 \rangle$ (which is finite in the $\delta\tau \rightarrow 0$ limit), the drift coefficient is a in Eq. (88), and there is a spatially varying diffusion coefficient

$$D(x) = 1 + e^x. \quad (92)$$

The exponential term in Eq. (92) is the effect of the nonlinearity γ in the tree problem.

Note that nonlinearity in the tree is unrelated to nonlinearity in the Fisher-KPP field $H(x, \tau)$ (which instead reflects branching of the tree).

H forms a traveling wave, whose speed c sets the exponential growth rate of Z [cf. Eqs. (68) and (72)]. On their own, the combination of ordinary diffusion and logistic growth (the Δ term) in Eq. (91) would give a traveling wave propagating to the *right* ($c > 0$), which corresponds to exponential growth of Z . Here, in one phase, this wave instead propagates backwards ($c < 0$): this is possible because of the drift term in Eq. (91). In the other phase, the wave attempts to propagate to the right but is stopped by the exponential growth of the diffusion constant at positive x , which prevents the buildup of H at large x . This results in $c = 0$.

There is therefore a transition between a phase where Z is exponentially small at large generation number and a phase where Z remains order 1. This is the toy model's version of the entanglement transition.

In the absence of the e^x term in $D(x)$, the velocity of a traveling wave with tail $H \sim e^{-\lambda x}$ is [98]

$$v(\lambda) = \lambda + a + \Delta\lambda^{-1} \quad (93)$$

[as we see by keeping only the order H terms in Eq. (91)], with a minimum at $\lambda_* = \sqrt{\Delta}$. Therefore, the linearized tree is in the glass regime [58] (Sec. G), where the traveling wave travels at speed $v_{\min} = v(\lambda_*)$, so long as $\Delta < 1$. This is the case we are interested in for the current circuit models, where $\lambda_* = 1/2$ at the entanglement transition.

The wave speed is then $c = 2\sqrt{\Delta} + a$, so in this toy model the analog of the entanglement transition is at $a_c = -2\sqrt{\Delta}$. Let us therefore write

$$a = -2\sqrt{\Delta} + \sigma. \quad (94)$$

We are interested in small positive σ , just inside the entangled phase. In principle we would like to solve for the stationary solution at late times,

$$\partial_x [D(x)\partial_x + 2\sqrt{\Delta} - \sigma]H + \Delta H(1 - H) = 0, \quad (95)$$

which we expect to exist when $\sigma > 0$. In the absence of a full solution, we consider the equation piecewise [109].

Let the position of the front, whose scaling with σ we wish to determine, be denoted $x_{\text{front}}(\sigma)$. We assume (and confirm below) that $x_{\text{front}}(\sigma)$ is large and negative at small σ .

First, at large positive x , the leading term in the equation is simply $\partial_x e^x \partial_x H = 0$, so the only solutions satisfying $H \rightarrow 0$ at large x have $H \sim e^{-x}$.

Second, consider $-|x_{\text{front}}(\sigma)| \ll x \ll 0$. In this regime we neglect both the variation of the diffusion coefficient and the $\mathcal{O}(H^2)$ term. From Eq. (93), we can find a stationary solution for positive σ only by making λ complex [109]. Keeping only the leading σ dependence,

$$H \sim e^{-\sqrt{\Delta}(x-x_0)} \sin(\phi + \Delta^{1/4} \sqrt{\sigma} x). \quad (96)$$

We would like to use the as-yet-undetermined constants x_0 and ϕ in order to allow this solution to match onto the solutions at large positive and negative x . Note that the slope of this solution on a logarithmic plot is

$$\partial_x \ln H = -\sqrt{\Delta} + \frac{\sqrt{\sigma} \Delta^{1/4}}{\tan(\phi + \Delta^{1/4} \sqrt{\sigma} x)}. \quad (97)$$

For generic x , this slope is close to $\sqrt{\Delta}$, because of the small factor $\sqrt{\sigma}$ in the second term. However, close to the zeroes of the tangent this is not true. This allows us to match on the right-hand side of the range [110], where the slope is steeper, so long as we take $\phi = \pi$ to leading order in σ .

Similar considerations on the left show the argument of the tangent must approach 0 as the vicinity of the front is approached. Therefore, to leading order in σ , the position of the front is

$$x_{\text{front}}(\sigma) = -\frac{\pi}{\Delta^{1/4}\sqrt{\sigma}}. \quad (98)$$

The constant x_0 in Eq. (96) then has the same leading term, to ensure that H is of order 1 in the front region.

Since $x_{\text{front}}(\sigma)$ also sets the average value of $\ln Z$,

$$Z^{\text{typ}} \sim \exp\left(-\frac{\pi}{\Delta^{1/4}\sqrt{\sigma}}\right). \quad (99)$$

By considering the tail of the distribution, we see that in the regime we are discussing, where $\Delta < 1$, the mean scales as

$$\langle Z \rangle \sim (Z^{\text{typ}})^{\sqrt{\Delta}} \sim \exp\left(-\frac{\pi\Delta^{1/4}}{\sqrt{\sigma}}\right). \quad (100)$$

Notice that Eq. (99) is

$$Z^{\text{typ}} \sim \exp\left(-\frac{\pi}{|\text{Im}\lambda_\sigma|}\right) \quad (101)$$

[also $\langle Z \rangle \sim (Z^{\text{typ}})^{\lambda_0}$, where λ_σ solves $v_\sigma(\lambda) = 0$. That is, it depends only on the function $v(\lambda)$ for the linear problem! Indeed the strength of the nonlinearity γ in Eq. (84) cannot appear, since it can be absorbed into a rescaling of Z (which does not affect $\ln Z$ at leading order).

This suggests that we can apply the result to the quantum tree of Sec. E, using Eq. (75) for v . This gives

$$Z^{\text{typ}} \simeq \exp\left(-\frac{C}{\sqrt{p_c - p}}\right), \quad (102)$$

with

$$C = \frac{\pi(1-p_c)}{\sqrt{2}} \sqrt{\left\langle \sum_i A_i^{1/2} (\ln A_i)^2 \right\rangle}. \quad (103)$$

For the Haar-random case C is given exactly by Eqs. (64) and (65). In terms of r [Eq. (43)],

$$Z^{\text{typ}} \simeq \exp\left(-\frac{1.482 \dots}{\sqrt{r_c - r}}\right). \quad (104)$$

Numerical results obtained by simulating the recursion relation in Eq. (84), and by solving the continuum traveling wave equation are shown in Fig. 15. These results indeed confirm the scaling of Z^{typ} close to the critical point.

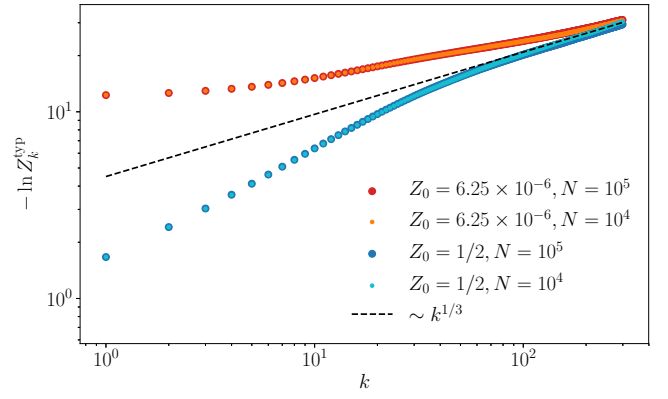


FIG. 16. The behavior of Z_k^{typ} with k at the critical point for two initial conditions, Z_0 , and two pool sizes, N . The data is consistent with a $k^{1/3}$ scaling of $\ln Z_k^{\text{typ}}$. Results are shown for the Haar ensemble and errorbars are smaller than the data points.

4. Tree entanglement at critical point

So far we discuss scaling in the two phases. Exactly at the transition, we might expect that Z_k tends to zero with k , but more slowly than in the distangled phase.

Figure 16 shows data for this for the Haar ensemble. The data is compatible with, though it does not clearly establish, the scaling $\ln Z_k^{\text{typ}} \sim k^{1/3}$, which is suggested by the following argument for the continuum model.

We expect that for the time-dependent equation x_{front} drifts sub-ballistically to the left. Let us conjecture that at a given time t , and in the range $x_{\text{front}} \ll x \ll 0$, the instantaneous solution of the nonlinear equation approximates sufficiently closely the traveling-wave solution $G^{(\lambda)}$ of the linear equation with the same instantaneous speed, $v = \dot{x}_{\text{front}}$. At the critical point ($\sigma = 0$), $v(\lambda)$ has a double zero at $\lambda = \sqrt{\Delta}$, so this means that $\lambda(t) \simeq \sqrt{\Delta} \pm i\Delta^{1/4}\sqrt{|\dot{x}_{\text{front}}|}$. This gives a solution like Eq. (96), but with $\sqrt{|\dot{x}_{\text{front}}|}$ in place of $\sqrt{\sigma}$. Equation (98) then becomes

$$x_{\text{front}} \sim -\frac{\pi}{\Delta^{1/4}|\dot{x}_{\text{front}}|^{1/2}}, \quad (105)$$

which gives $x_{\text{front}} \sim (3\pi^2 t / \sqrt{\Delta})^{1/3}$. These values for the exponent and the prefactor are in good agreement with a numerical solution of Eq. (91) at $a = -2\sqrt{\Delta}$ (we check the case $\Delta = 1/4$).

If σ is small but positive there must be a crossover at a large time t_{sat} from $x_{\text{front}} \sim -t^{1/3}$ to $x_{\text{front}} \sim -1/\sqrt{\sigma}$. This suggests $t_{\text{sat}} \sim \sigma^{-3/2}$, which also agrees well with numerical solutions.

I. Quantum trees: other universality classes?

Above we noted that *a priori* there are two possibilities according to whether the entanglement transition takes place within the glass or the paramagnetic regime

of the linearized recursion relation: $\lambda_* < 1$ and $\lambda_* > 1$, respectively.

However, our approach required the statistical invariance of the node tensor $t_{b_1, b_2, b_3, \dots, b_\ell}^a$ under $U(2)$ rotations on a leg. (We are free to allow for an arbitrary branching number ℓ .) This invariance is necessary so that we can write a recursion relation for singular values only: otherwise we need a combined recursion relation for singular values and singular vectors. For any such tree, the argument of Sec. IV F and Appendix 2 shows that $\lambda_* = 1/2$ at the transition. That is, the recursion relation is of the form [111]

$$Z_{k+1} = \sum_{i=1}^{\ell} A_i Z_k^{(i)} + \mathcal{O}(Z^2), \quad (106)$$

with $\langle A_i \rangle = 1$ and $\langle A_i^{1/2} \ln A_i \rangle = 0$, which is sufficient to ensure $\lambda_* = 1/2$ at the critical point (Sec. IV F).

In this class of trees the weak correlations between the top and base in the disentangled phase are dominated by only a subgraph of the tensor network that contains a few paths from top to bottom. For this broad class of trees we expect the universal scaling described above. Therefore, within the class of trees that our formalism applies to there is no freedom to vary λ_* .

However, it is interesting to ask what happens in trees where the unitary invariance property is broken. Breaking this invariance introduces correlations between singular values and singular vectors. A plausible guess, at least if these correlations are not too strong, is that in this setting the same toy model nevertheless captures the universal scaling. If this is the case (which we do not determine here) then the next question is whether in these more general models it is possible to vary the critical value of λ_* away from $1/2$.

With this somewhat speculative motivation (and for completeness), below we extend the analysis of critical scaling in the toy model to the regime $\lambda_* > 1$.

Our analysis is also restricted to trees with bond dimension 2. A recursion relation (for the subleading squared singular values squared) may be formulated for trees with larger bond dimension, but has a more complicated structure, even at lowest order. It would be interesting to study this further.

1. Scaling of nonlinear recursion: $\Delta > 1$

We return to the toy model of Sec. 2 in the continuum limit, now with $\Delta > 1$. The near-critical regime inside the entangled phase is now

$$a = -(\Delta + 1) + \sigma, \quad (107)$$

with $0 < \sigma \ll 1$ representing the control parameter that drives the entanglement transition.

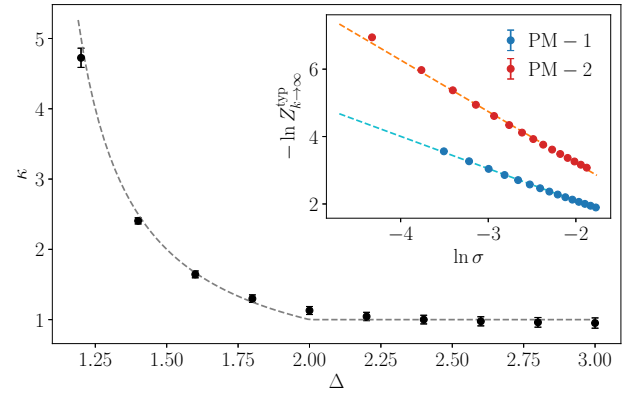


FIG. 17. Main panel: the exponent κ that defines the critical divergence of Z^{yp} [see Eq. (108)] as a function of the parameter Δ in the power-law regime $\Delta > 1$. Black points show results from a numeric solution of Eq. (91) whereas the gray dashed line shows $\kappa = \max\{1/(\Delta - 1), 1\}$. Details of the numeric solution and definition of error bars are given in Appendix 3. Inset: the critical divergence of Z^{yp} from the discrete toy tree for two sets of parameters in the paramagnetic phase, PM-1 and PM-2, see text for details. For PM-1, the expected exponent is -1 and the best fit exponent is -0.96 whereas for PM-2, the expected and the fitted exponents are -1.97 and -1.53 , respectively.

The difference from the case studied above [cf. Eq. (96)] is that the solutions λ of $v_\sigma(\lambda) = 0$ are no longer complex: instead there is a real solution at $\lambda = 1 + \mathcal{O}(\sigma)$, and a larger real solution at $\lambda_+ = \Delta + \mathcal{O}(\sigma)$. That is, if we neglect both the nonlinearity in H and the x dependence of the diffusion constant, the stationary solution is a sum of two exponentials, in contrast to Eq. (96).

In Appendix 3 we study this regime via the equation for $R = \partial_x \ln H$, which interpolates between 0 for $x \ll x_{\text{front}}$ and -1 for $x \gg 0$. We conclude that

$$Z^{\text{yp}} \sim \sigma^\kappa, \quad \kappa \equiv \max\left\{\frac{1}{\lambda_+ - 1}, 1\right\}, \quad (108)$$

where the denominator appearing in κ is the difference of the two solutions to $v(\lambda) = 0$ at the critical point $\sigma = 0$. In the present model, $\kappa = \max\{1/(\Delta - 1), 1\}$, but we conjecture that the form in Eq. (108), which requires only knowledge of the speed function $v(\lambda)$ of the linearized problem, applies to a wider set of models.

Our argument in Appendix 3 is not rigorous, so we compare the formula $\kappa = \max\{1/(\Delta - 1), 1\}$ with a numerical solution of the continuum equation. Results are shown in Fig. 17 and are in fairly good agreement with the prediction.

Numerical solution of the continuum equation suggests that the above exponent κ also determines the decay of Z right at r_c ,

$$Z \sim t^{-\kappa}. \quad (109)$$

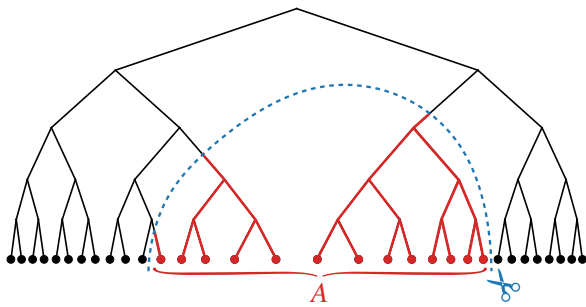


FIG. 18. Schematic: tree tensor-network wave function for a chain of spins, with a subset A of spins and the corresponding minimal cut indicated. We propose a modified minimal cut formula for the entanglement.

We also study the discrete toy tree model in Eq. (84) numerically in the regime with $\lambda_* > 1$. We find polynomial scaling of Z^{typ} near the critical point as expected. The numerical estimates of the exponents differ somewhat from the predicted ones, which we attribute to finite size limitations. See Fig. 17 (inset) for examples. The parameters corresponding to PM – 1 are $p_1 = 0.15$, $p_2 = 0.85$, $\epsilon = 0.95$, and $\gamma = 0.5$, as such $\Delta = 3.97391$ and we expect $Z_{k \rightarrow \infty}^{\text{typ}} \sim \sigma^{-1}$. Indeed the best-fit exponent from our simulation is $\kappa = -0.96$. On the other hand, the parameters corresponding to PM – 2 are $p_2 = 1$, $\epsilon = 1.5$, and $\gamma = 0.5$ such that $\Delta = 1.5066$ as such the expected exponent is $\kappa = -1.97$. We however find a best-fit exponent of $\kappa = -1.53$ and attribute the discrepancy to finite size of the pool and distance from the critical point.

J. Trees, entanglement, and min cut

So far we characterize the entanglement between the top of the tree and the base. We now apply this to more general entanglement quantities in the tree.

Figure 18 is a schematic of a wave function for a chain of spins that is given by a tree tensor network (note that there is no longer a free bond at the top). Here we consider the entanglement $S(R)$ of a set A of $R \gg 1$ contiguous spins in a much larger chain. This problem has also been tackled recently in Ref. [68] using a different method: see Sec. IID.

As is well known, in such a geometry the minimal cut sketch suggests the scaling $S \sim \ln R$ [67–69], which is the number of bonds cut for “typical” choices of the placement of the region A (the tree strongly breaks translational invariance). Figure 18 shows an example of a minimal cut in a small tree. The logarithmic scaling is presumably correct in the entangled phase, but what happens close to the transition? For simplicity we consider the second Rényi entropy.

Note that the minimal cut in Fig. 18 lops off a disjoint set of smaller subtrees, marked in red and thick. We assume that the region is placed so that this is the case. In this setting, a natural conjecture for the tree is that the

universal scaling forms for the entanglement close to the transition and in the disentangled phase are given by a “modified minimal cut” formula: we first find the geometrical minimal cut, but then *weight* the contribution to the entanglement of a bond at height k (k generations above the base) by an amount that depends on Z_k . Since we are interested in the region close to the critical point and large k , we assume $Z_k \ll 1$.

What should this weight be? The simplest case is where the minimal cut breaks only one bond, i.e., where region A corresponds to a single connected subtree. In such cases the minimal cut breaks the full tree into two subtrees, and each one is characterized by a Z value. (In general one of them has an irregular structure, with different numbers of generations for different branches, but we can still use the recursion relation to compute its Z value.) A 2×2 matrix calculation shows that the second Rényi entropy S_2 is proportional to the product of these Z values, $S_2 \propto ZZ'$, to leading order [112]. Using the fact that Z_k^{typ} is asymptotically nonincreasing in k , the Z values of the subtrees are both of typical size Z_k^{typ} for k equal to the height of the cut bond.

This suggests the conjecture

$$S_2(R) \sim \sum_{k=1}^{c \ln R} Z_k Z'_k. \quad (110)$$

In this schematic formula, $c \ln R$ is the maximum height reached by the minimal cut, and Z_k, Z'_k are random variables. All order-1 constants have been neglected, since we aim only to capture the asymptotic scaling with R and with the distance from the critical point. We can confirm Eq. (110) explicitly in an artificial limit in which the scale of the Z s tends to zero, with R arbitrary but fixed: this is described in Appendix 4. However, in the physical problem we wish to take R to infinity, so this does not prove the conjecture.

Consider the first class of trees (including those with the statistical invariance property of the node tensors, for example those appearing in the Haar circuit). Let r be an arbitrary parameter that drives the tree’s entanglement transition. The results in the previous sections and Eq. (110) yield

$$\overline{S_2(R)} \sim \exp\left(-\frac{\text{const}}{\sqrt{r_c - r}}\right) \ln R \quad (r \lesssim r_c), \quad (111)$$

$$\overline{S_2(R)} \sim \mathcal{O}(1) \quad (r = r_c). \quad (112)$$

Surprisingly, the entanglement is order 1 at large R at the critical point, because of the rapid decay of Z with k (Sec. 4). This is also true in the disentangled phase.

In the previous section we speculate about the existence of trees with an effective value of $\Delta > 1$. If such

trees exist, then the same reasoning as above gives $S_2(R) \sim (r_c - r)^{2\kappa}$ for $r \lesssim r_c$, with a variable exponent $\kappa = \max\{\Delta - 1, 1\}$. The entanglement right at r_c is again $\mathcal{O}(1)$.

K. Connecting back to the quantum circuit

Our original motivation for studying the tree is the conjecture that, for the forced-measurement circuit models described in detail in Sec. 3, the critical point r_c of the appropriate tree ensemble is also the critical point for the circuit.

Here we give an argument that bounds the operator entanglement in the circuit in terms of the entanglement in the tree. This argument is very heuristic: a task for the future is to make the connection between the circuit and the tree more precise.

The basic idea is to imagine breaking a bond in the interior of the FMPT circuit, and to ask how much effect this can have on properties of the nonunitary time-evolution operator V . Let b be a bond inside the circuit at time coordinate approximately $t/2$. Then V_b is a modification of V in which bond b is broken.

Starting from b we imagine marking the two trees T and T' attached to either end of it, using the convention in Sec. 4, where bonds with projectors on them are removed. We stop after k generations, choosing the largest possible k such that these are indeed two disjoint trees (no loops). Therefore, k should be of order $\ln N$. We assume that the number of spins N is very large, so that the typical size of Z and Z' for these trees (the minimal singular value squared) is given by the asymptotic large k result. Close to r_c , this typical value is small.

Together T and T' , connected by b , form a tensor network \tilde{T} . This can be seen as a state in a tensor product Hilbert space $\mathcal{H} \otimes \mathcal{H}'$ associated with the bonds on the boundary of T and T' , respectively. We may form the corresponding singular value decomposition of \tilde{T} : the smaller of its two singular values is of order $\sqrt{ZZ'}$, in terms of the Z values or T and T' (assumed small, since we are in the critical regime).

Breaking the bond is defined to mean dropping this minimal singular value. *Formally*, this induces an error that is of order $\sqrt{ZZ'}$ in Z and Z' . After averaging over the local unitaries, the error in any physical quantity is—again, formally—of order ZZ' [113]. This suggests that the average change in S_2 when we break a single bond is *at most* of order $\langle Z_\infty \rangle^2$. This is of course far from being a proof, because in principle the small term ZZ' in the formal expansion could be systematically compensated by a large prefactor.

Assuming this bound, we may straightforwardly bound the plateau value $s(r)N$ of the operator entanglement close to r_c (Sec. II, Sec. I). Since breaking all the N bonds in a timeslice reduces the entanglement to zero, we must have

(c_1 and c_2 are constants):

$$s(r) \leq c_1 \exp\left(-\frac{c_2}{\sqrt{r_c - r}}\right) \quad (113)$$

for small $r_c - r$. Assuming also our conjecture that r_c is the same for the tree and the circuit, this indicates that $s(r)$ vanishes extremely rapidly as the critical point is approached. In turn, $s(r)$ is related to the scaling of the exponential timescale in the entangled phase (Sec. I), just as it is in the classical problem (Sec. III).

We note that the bound Eq. (113) on the scaling need not be tight. This can be understood by considering the analogous argument for the classical minimal cut problem.

Above, our bound used $\langle Z \rangle$. The analogous quantity in the classical problem is the probability that a given tree is infinite. This is essentially the order parameter in the classical problem, scaling like $f_\infty \sim (r_c^{\text{cl}} - r)$. We can bound the cost of the classical minimal cut S_0 as follows. Consider all the bonds of the percolation configuration that traverse some timeslice, say at time $t/2$. Each bond has a probability approximately f_∞^2 that the two trees attached to either end of it are *both* infinite. These are the only bonds we need to cut (the others lie in disconnected clusters or dangling ends). This shows that $s_{\text{cl}}(r)$ goes to zero *at least* as fast as $(r_c^{\text{cl}} - r)^2$ close to the transition.

This bound is consistent with, but weaker than, what we argue is the true scaling in the classical problem, $s_{\text{cl}}(r) \sim (r_c^{\text{cl}} - r)^{5/2}$ (Sec. III D).

V. SIMULATIONS OF QUANTUM CIRCUITS

Having made a connection between trees and all-to-all circuits, we now turn to the numerical simulation of the latter. Exponentially large in system size Hilbert-space dimensions restrict us to systems with $N \leq 20$ spin 1/2. We simulate both measurement circuits and forced-measurement circuits, keeping in mind that the results obtained from the tree apply only to the latter. Unless specified, the results shown here are for the Haar ensemble Eq. (37) [we also comment briefly on the $\Delta t = 0.3$ ensemble, Eq. (38)].

All-to-all circuits have no spatial structure. Consequently, the entanglement transition does not entail a volume-to-area law transition in the entanglement associated with a spatial bipartition of a state. Instead, we consider two observables, which quantify the amount of quantum information transmitted from the initial to the final time: (i) the time evolution of the *operator entanglement entropies* (opEE) of the nonunitary evolution operator V , and (ii) the overlap of two initially orthogonal states that are both evolved using V .

We show that the entanglement transition separates an entangled phase at $r < r_c$, wherein an extensive amount of quantum information is retained for an exponentially long

time, from a disentangled phase at $r > r_c$ wherein memory of the initial state is rapidly lost. This is in agreement with analytical results for the quantum problem in Sec. I, and is qualitatively similar to what we find in the classical toy model in Sec. III.

First, we give evidence for a plateau in the operator entanglement for r below a critical value. We then turn to observable (ii) above: since this does not require exact diagonalization of V , it allows larger N to be accessed. We use this observable to define a timescale $\tau(r, N)$, and show that this timescale scales exponentially with N inside the entangled phase. Details of numerical calculations are relegated to Appendix D.

A. Operator entanglement

The amount of information carried from the bottom of the circuit to the top can be quantified via the OEE of V . In the case of measurements, where we must choose an initial state in order to define the Born-rule probabilities, we take this state to be a product state (with spins aligned in the positive x direction, $|\rightarrow\rightarrow\cdots\rightarrow\rangle$).

The OEE is obtained from the singular value decomposition

$$V = \sum_{j=1}^{\mathcal{D}_H} \mu_j |j\rangle_t \langle j|_0, \quad (114)$$

where $\mathcal{D}_H = 2^N$ is the Hilbert-space dimension, and $\{|j\rangle_0\}$ and $\{|j\rangle_t\}$ are bases corresponding to the initial and final time. (We leave the t dependence of V implicit.) The OEE is

$$S_n = \frac{1}{1-n} \ln \sum_{j=1}^{\mathcal{D}_H} \lambda_j^{2n}, \quad (115)$$

where $\lambda_j \equiv \mu_j / \sqrt{\sum_j \mu_j^2}$. For a unitary V , S_n takes on its maximal value of $N \ln 2$. Any reduction compared to this value reflects loss of information between initial and final time due to worldlines of the spins broken by measurements. S_n is bounded from above by the minimal cut separating the initial and final times. Close to the FMPT transition, we also have the conjectural bound Eq. (113) on the scaling form for $s_2(r) = S_2/N$ in the plateau region.

Results for S_1 for the Haar ensemble with forced measurements are shown in Fig. 19. In the top panel, we plot the entanglement density S_1/N vs t for $r = 0.1$ and various systems sizes. After an initial linear decrease with t associated with the first measurements, there is a time regime where S_1/N increases with N . This suggests the emergence of a plateau in S_1/N at large N (Sec. II).

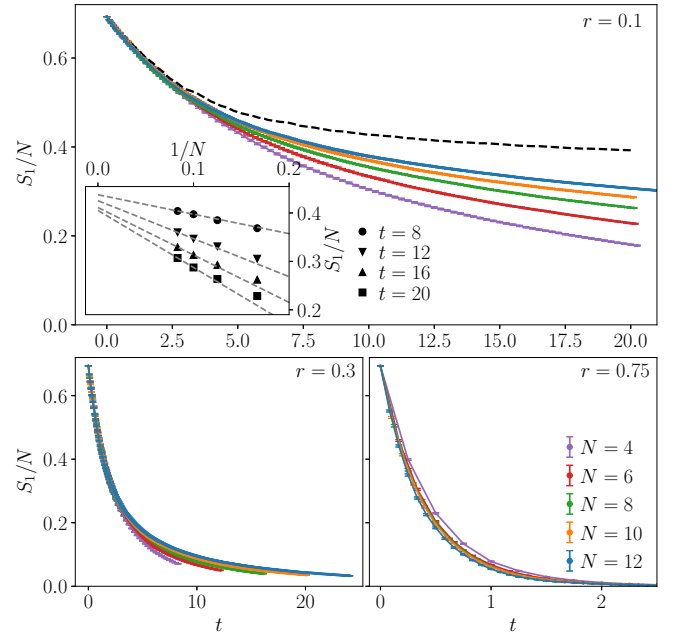


FIG. 19. The OEE, S_1 , of the nonunitary time-evolution operator [see Eq. (115)] for the Haar ensemble with forced measurements. The top panel is for $r = 0.1$, deep in the entangled phase where there is a plateau in S_1/N in the $N \rightarrow \infty$ limit. The extrapolated $N \rightarrow \infty$ value is shown as the black dashed line; the extrapolation in N is shown for a few exemplary time points in the inset. The bottom panels correspond to $r = 0.3$ and $r = 0.75$, the latter being the putative r_c obtained from the Haar tree. Note that the value of S_1/N is already quite small in the entangled phase at $r = 0.3$. All data is averaged over 5000 realizations.

Recall that in the entangled phase we expect a nonzero value for

$$s_n(r) \equiv \lim_{t \rightarrow \infty} \lim_{N \rightarrow \infty} \frac{S_n(N, t, r)}{N}, \quad (116)$$

and that when N is finite but large, S_1/N remains close to $s_1(r)$ over a range of times that grows exponentially with N .

To give evidence for the nonzero value of $s_1(r)$ at $r = 0.1$, we extrapolate the data to $N = \infty$ for each value of the time [114]. This $N \rightarrow \infty$ extrapolation is shown as a dashed line in the figure. It is consistent with a plateau, extending to $t = \infty$, with $s_1(0.1) > 0$. (We defer an analysis of timescales to the following subsection.) A similar plateau was observed in Ref. [12] in Clifford circuits.

It is clear that the plateau value $s(r)$ decreases very rapidly with increasing r . In the lower left panel we plot S_1/N for the same set of N and $r = 0.3$. This r value is still far from the conjectured location of the critical point obtained from the tree ($r_c \simeq 0.749$). An increase of S_1/N with N is still observed, but it is clear that s_1 (assuming it is nonzero) is small. It is tempting to associate this with the exponential scaling in Eq. (113), which suggests that $s(r)$

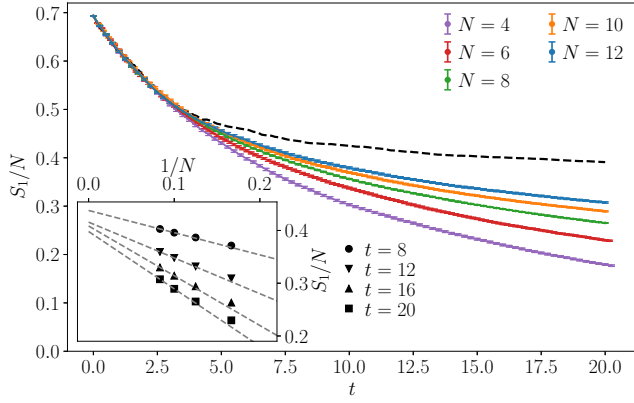


FIG. 20. As in Fig. 19(top), the OEE density, S_1/N , of the nonunitary time-evolution operator for the Haar ensemble but with measurements for $r = 0.1$. The data is consistent with there being a plateau in S_1/N in the $N \rightarrow \infty$ and $t \rightarrow \infty$ limit. All data is averaged over 5000 realizations.

goes to zero very fast as the critical point is approached, so that a plot of $s(r)$ against r would be very flat for $r \lesssim r_c$.

On the other hand, at $r = 0.75$ (lower right panel), S_1/N decays exponentially to zero, with a very weak N dependence and no indication of saturation at large t . In fact, the trend with increasing N is in the opposite direction to the cases $r = 0.1$ or 0.3 .

Thus, the OEE for these system sizes is consistent with an entanglement transition, occurring below the classical critical point, and with the expected plateau for S/N in the entangled phase. However, the rapid decay of the plateau value $s(r)$, and the weak N dependence, make it hard to pin down the position of the transition. We check (but do not show) that the data for S_2 is qualitatively similar to that for S_1 .

We find the same qualitative features for the Haar circuit with true measurements. Figure 20 shows the case $r = 0.1$. In fact at this relatively small value of r , the data for forced measurements and measurements is almost indistinguishable.

B. State overlap and timescales

We now turn to the overlap of two initially orthogonal states undergoing time evolution with the nonunitary operator V :

$$O(t) \equiv |\langle \psi^{(2)}(t) | \psi^{(1)}(t) \rangle|^2. \quad (117)$$

We also define the “distance” $\mathcal{D}(t)$ as

$$\mathcal{D}(t) = 1 - O(t). \quad (118)$$

The states are initiated as product states in the σ^x basis, $|\psi^{(1)}(0)\rangle = |\rightarrow \rightarrow \dots \rightarrow\rangle$ and $|\psi^{(2)}(0)\rangle = |\leftarrow \leftarrow \dots \leftarrow\rangle$

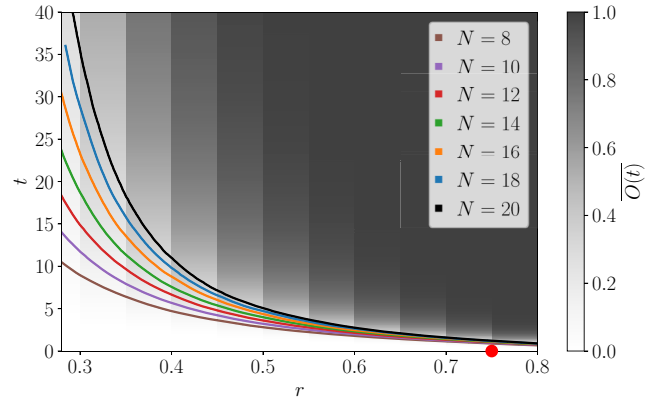


FIG. 21. Heatmap: the squared overlap $\overline{O(t)}$ between the states $|\psi^{(1)}(t)\rangle$ and $|\psi^{(2)}(t)\rangle$ [see Eq. (117)], as a function of time t and measurement rate r for $N = 20$. The different lines show the contours for $\overline{O(t)} = 1/2$ for different N . The data corresponds to the Haar circuit with forced measurements. The red mark at $r = 0.75$ denotes the critical point from the Haar tree.

and are evolved using the nonunitary operator

$$|\psi^{(j)}(t)\rangle = \frac{V^t |\psi^{(j)}(0)\rangle}{\sqrt{\langle \psi^{(j)}(0) | V^\dagger V | \psi^{(j)}(0) \rangle}}. \quad (119)$$

In the case of forced measurements the spins are always projected along the positive σ^z direction, and there is a symmetry between $|\psi^{(1)}(t)\rangle$ and $|\psi^{(2)}(t)\rangle$. In the case of measurements we use $|\psi^{(1)}(t)\rangle$ to determine the Born-rule probabilities, so this symmetry is absent.

$O(t)$ is another way to quantify the amount of information retained from the initial state. In the limit $r = 0$, where V is unitary, the two states remain orthogonal for all time, $O(t) = 0$. In the opposite limit $r = 1$, where no unitaries are applied, $O(t)$ will be exactly one as soon as all spins are measured. For any fixed $r > 0$, and for a fixed value of N , the states will inevitably converge to 1 as $t \rightarrow \infty$, because they are being subjected to the same projections. However, we expect the timescale for this to grow exponentially with N in the entangled phase.

For a broadbrush view, we first show $O(t)$ as a heatmap in the space of r and t for the Haar forced-measurement circuit with $N = 20$ spins: Fig. 21. $O(t)$ grows towards unity with an r -dependent timescale. One way to define a timescale is using the contour $\overline{O(t)} = 1/2$. This is shown not only for $N = 20$, but also for smaller values of N , in the figure. The conjectured r value of the phase transition is marked by the red dot. The timescale grows rapidly as r is decreased. It also shows a clear N dependence in the entangling phase, which becomes much weaker on approaching the transition point.

Figure 22 shows the time dependence of the overlap in more detail for $r = 0.5$ and $r = 0.75$ [115]. In both cases the time required to achieve a given value of O increases

with N , but in the latter case this is mostly a shift of the curve, whereas in the former case there is the clear sign of an increasing time constant for the exponential approach of O to 1.

We use the exponential approach of the overlap to unity to define a timescale $\tau(r, N)$. Since at late times $\mathcal{D} = 1 - O$ is exponentially small, and may have a broad distribution, we choose to look at its typical value. We define this by $\ln \mathcal{D}^{\text{typ}}(t) \equiv \overline{\ln \mathcal{D}(t)}$, where instances in which $\mathcal{D}(t)$ is exactly zero are excluded from the average (similar to the treatment of the singular value Z in the quantum tree, Sec. D). At late times this shows an exponential decay,

$$\ln \mathcal{D}^{\text{typ}} \sim -\frac{t}{\tau(r, N)}. \quad (120)$$

Data for $\ln \mathcal{D}^{\text{typ}}(t)$ vs t are shown in Fig. 23, for the same values of N and r as in Fig. 22. We see clear exponential decay. In fact, Fig. 23 vividly shows the qualitative difference between the cases of $r = 0.5$ and 0.75 : while τ grows with N for $r = 0.5$, it appears essentially N independent for $r = 0.75$. Data for the circuit with measurements (not shown) is qualitatively similar.

We now analyze $\tau(r, N)$ in the entangled phase. This is the asymptotic slope of plots like Fig. 23. We extract this from a plot of $\tau_{\text{eff}}(t) = -(\frac{d \ln \mathcal{D}^{\text{typ}}}{dt})^{-1}$, the time-dependent slope: at late times, $\tau_{\text{eff}}(t)$ should stabilize at the value τ . Representative data for $\tau_{\text{eff}}(t)$ and the plateaux therein are shown in Appendix D, see Figs. 39 and 40.

It turns out that finite-time effects become significant at larger values of r , but for r not too large we are able to obtain an estimate of τ . The data (shown in Appendix D) is consistent at small r with exponential-in- N growth of the timescale:

$$\ln \tau(r, N) \sim a(r)N. \quad (121)$$

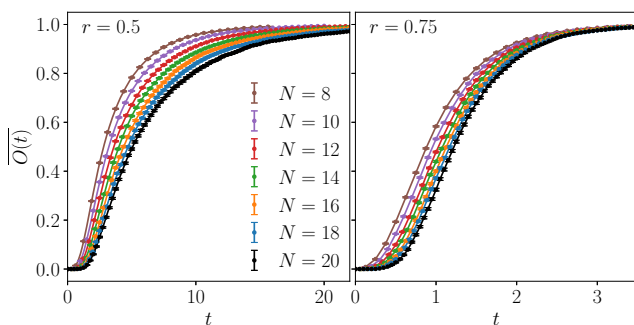


FIG. 22. The squared overlap between the states $\overline{O(t)}$ as a function of time t for two exemplary values of r for different N . The data corresponds to the Haar circuit with forced measurements. Note the difference in the range of times shown in the two panels. It is also possible to see the emergence of a plateau at $\overline{O(t)} = 0$ at early times.

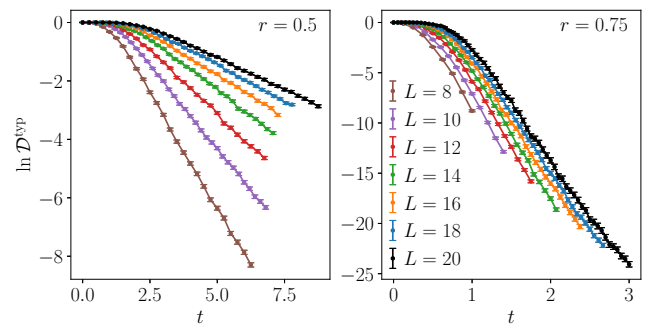


FIG. 23. The typical distance $\mathcal{D}(t) = 1 - O(t)$ between the states $|\psi^{(1)}(t)\rangle$ and $|\psi^{(2)}(t)\rangle$ as a function of time t for the same parameters as in Fig. 22, showing exponential convergence of the two states at late time. Note the difference in the scales shown in the two panels.

The coefficient $a(r)$ is plotted against r in Fig. 24. This figure also shows data for the case of true measurements.

We expect $a(r)$ to vanish at the critical point with $a(r) \sim s(r)$ (see Sec. I). Unfortunately, hamstrung by severe finite-size effects, we are not able to estimate the critical point accurately [116]. The data is certainly consistent with a critical point for the FMPT, which is below the conjectured value $\simeq 0.749$. However, we speculate that this is instead a symptom of $a(r)$ vanishing very rapidly as r_c is approached, as is suggested by the essential singularity in Eq. (113). The dashed line in the figure shows this exponential form with $c_1 = 39.4$ and $c_2 = 3.8$. These values have no theoretical significance: this line is simply to indicate the possibility of $a(r)$ remaining very small even for r considerably below r_c .

The data for S_1 at small values of r is very close for measurements and forced measurements, as noted above.

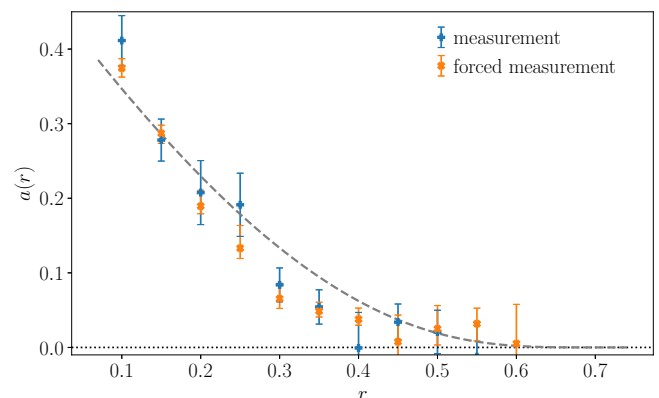


FIG. 24. The coefficient $a(r)$ in the exponential dependence of the timescale $\tau \sim e^{a(r)N}$ on N , as a function of r for the Haar circuit with measurements and that with forced measurements. The dashed line is a guide to eye for a function $c_1 e^{-c_2/\sqrt{r_c-r}}$ with $r_c = 0.75$, $c_1 = 39.4$ and $c_2 = 3.8$. See Fig. 41 for the fits used to extract $a(r)$.

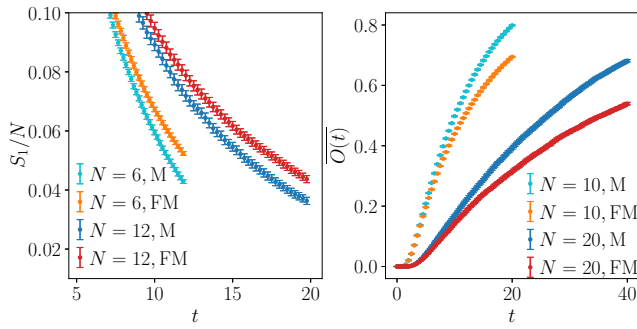


FIG. 25. Evidence that the forced-measurement (FM) protocol leads to higher entanglement than the measurement (M) protocol at the same value of r . The data is for $r = 0.3$. The left panel shows S_1/N whereas the right panel shows the average overlap $\overline{O}(t)$.

We do see differences between the two cases at intermediate r , with the forced-measurement circuit having slightly larger entanglement at a given r . This is shown for S_1/N in Fig. 25(left). The comparison between the overlap data for the two cases at the same value of r [Fig. 25(right)], is also consistent with the above, with $\overline{O}(t)$ growing more slowly in the forced-measurement case. This hints that r_c for the measurement case may be lower than that for the forced-measurement case, but our data does not allow us to determine this.

While the data above is shown for the Haar circuit, we perform the same set of numerical calculations for the $\Delta t = 0.3$ circuit as well. The results are qualitatively similar. Consistent with the results from the quantum tree, r_c appears to be smaller for the $\Delta t = 0.3$ circuit compared to the Haar.

VI. FIELD THEORIES FOR MEASUREMENT AND ENTANGLEMENT TRANSITIONS

A key question about the MPT, not previously resolved, is whether there is a simple Landau-Ginsburg-Wilson-like field theory that captures its universal properties. This question is also unresolved for entanglement transitions in random tensor networks (RTNs), and for the closely related FMPT. In this section we propose candidates for these field theories. (In this section the space-time dimensionality $D = d + 1$ is allowed to be arbitrary.) We obtain two Lagrangians, one for the MPT, and one for both FMPT and RTN. Surprisingly, these two Lagrangians are quite different in their structure, having for example different values for the upper critical dimension.

Microscopically, random circuits and random tensor networks can be mapped to lattice statistical mechanics models [13,14,34–36,42,44,45]. These are effective spin models where the “spin” is a group element in the permutation group S_N for N objects (we review this below; here N is a replica number and not the number of qubits as

in previous sections). However, using these lattice models to guess appropriate continuum field theories is nontrivial for various reasons, one of them being a replica limit that is necessary to handle randomness. “Replica” lattice models were described for a random tensor network in Ref. [36], for Haar circuits in Ref. [34], and for circuits with measurement in Refs. [13,14].

Previous work pointed out that in certain limits (either by artificially deforming the weights in the effective spin model [36], or by taking a $q \rightarrow \infty$ limit in the measurement problem [13,14]) one could access a fine-tuned point where the effective spin model had a simple continuum theory, namely that of percolation. While this was a useful step, this fine-tuned point has an infinite number of relevant perturbations [36] so unfortunately this does not provide a definite Lagrangian for the physical phase transitions of interest. Another approach is to study Ising models that are obtained by simply omitting the replica limit, roughly in the spirit of an annealed average in conventional disordered systems [24,27,35]. These are useful toy models for various phenomena in the entangled phase [24,27] (we give an explanation for why this is, building on Ref. [43]) but they cannot capture the correct critical properties. Therefore, we attempt here to formulate explicit continuum replica field theories.

We emphasize that these theories are speculative conjectures, based on writing down the simplest Lagrangians compatible with the basic symmetries of the problem. It is certainly possible that in fact something more complicated happens in the continuum. Indeed, the exponential scaling we find in the tree seems to mean that it is not described by the high-dimensional limit of the field theory for RTNs proposed below (see Sec. H). How to resolve this tension is a question for the future.

We first review the replica approach, the inevitable global symmetry of the field theories we are looking for, and the emergence of permutations in the simplest Haar-random models (Secs. IV A and IV B are largely review). We then discuss the coarse graining of these degrees of freedom (Sec. C). Next we note that these degrees of freedom have a more general meaning in terms of Feynman trajectories in the circuit [43,117]. This picture motivates an alternative derivation of a lattice field theory, which in turn suggests a simpler continuum formulation (Sec. VID). Our discussion also suggests an alternative way of thinking about the effective statistical mechanics of random tensor networks, in a way that is closer to traditional replica formulations of random magnets.

Then we discuss the issue of “replica group theory” for the MPT on one hand, and the RTN and FMPT on the other: that is, constraints on the field theories associated with the replica symmetry [118]. We propose the simplest candidate Lagrangians in each case (Secs. E–G). We discuss some of the basic consequences of the simpler of these Lagrangians, that for the MPT (Sec. H). Our discussion of

these field theories is relatively schematic: further details will be given in Ref. [119].

Section I, which is independent of the field theories proposed here, addresses scaling within the two phases, not necessarily near the critical point. Finally Sec. J describes variations of the measurement problem that are in distinct universality classes, for example, models with free-fermion structure or with additional symmetries.

A. Multilayer circuits and replica symmetry

The crucial symmetries of the problem arise when dynamical quantities are written in terms of a multilayer circuit, illustrated schematically in Fig. 26. (We use the language of a circuit, with d spatial dimensions and one time dimension, but analogous considerations apply to a $D = d + 1$ dimensional RTN.) This multilayer circuit is a discrete analog of a path integral with multiple forward and backward paths, and it arises when we write powers of the reduced density matrix, say for the final state, in terms of the circuit. Let us briefly review this.

The layers are N identical copies of the original circuit $V(t)$ and N copies of its complex conjugate $V(t)^*$. We call these “forward” and “backward” layers, respectively. Formally, the multilayer circuit with a given N may be written

$$V^{(N)} \equiv \underbrace{V \otimes \dots \otimes V}_N \otimes \underbrace{V^* \otimes \dots \otimes V^*}_N. \quad (122)$$

The physical quantity of interest will dictate the boundary conditions at the top and bottom: for example, contractions of indices between layers, or contraction of the bond indices at the bottom of a layer with an initial wave function. We review this in a simple setting in Sec. IV A. An important feature is the replica trick: N must be left free at intermediate stages of any calculation and then sent to

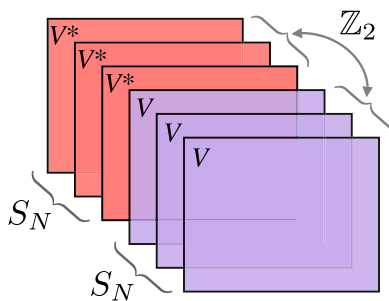


FIG. 26. Schematic: multilayer circuit (or tensor network) with N copies of the circuit V and N copies of its complex conjugate V^* . Each layer of the circuit has free bond indices at the bottom and top (initial and final time), which are not shown. The actions of the left and right permutation symmetries and the \mathbb{Z}_2 exchange that make up the internal symmetry group $G_N = (S_N \times S_N) \rtimes \mathbb{Z}_2$ of the replica field theories are illustrated.

a limiting value at the end [13,14,34,36]. (A special case where replicas can be omitted is mentioned in Sec. IV B below.) Replicas allow us to handle denominators that arise because of the normalization of states [36] [nonunitarity means these normalization factors are nontrivial, see Eq. (3)] and/or to deal with logarithms in the definition of the entanglement entropies. See Refs. [13,14,36] for detailed discussions of this for the RTN and MPT.

The global symmetry of the effective models arises ultimately from a simple invariance of $V^{(N)}$ under various operations. $V^{(N)}$ is clearly invariant under (i) permutations of the forward layers among themselves; (ii) permutations of the backward layers among themselves; and (iii) complex conjugation accompanied by exchange of all the forward layers with all the backward layers [43]. Together these make up the symmetry group:

$$G_N \equiv (S_N \times S_N) \rtimes \mathbb{Z}_2. \quad (123)$$

Here the \mathbb{Z}_2 is generated by (iii) above. G_N is a symmetry of the bulk structure of the tensor network; it will in general be broken by boundary conditions, e.g., by a choice of index contractions at the boundary of $V^{(N)}$.

A formal way to see the importance of this symmetry is via explicit mappings of random circuits or random tensor networks onto effective lattice spin models. We review this next. We give an alternative picture below in Sec. VID, by introducing an Edwards-Anderson-like field in a multilayer tensor network (this alternative picture may be more intuitive for those familiar with random magnets).

In simple models, averaging over the random tensors or unitaries leads to effective lattice magnets in which the “spins” σ (not to be confused with the physical spins that the circuit acts on) are valued in the permutation group [13,14,34–36,42,44,45]:

$$\sigma \in S_N. \quad (124)$$

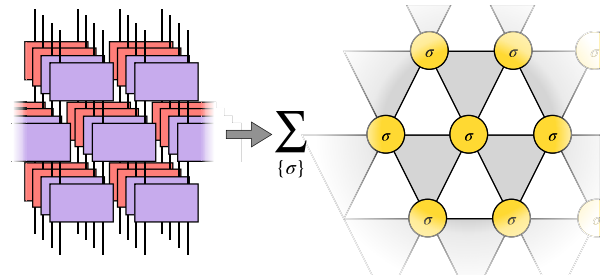


FIG. 27. Schematic: mapping a circuit with unitaries and possibly measurements to an effective spin model. Each physical unitary gives rise to a separate spin degree of freedom $\sigma \in S_N$ (yellow circles). These spins have interactions on downward pointing triangles (shaded). We refer to σ and its continuum versions as the “pairing field.”

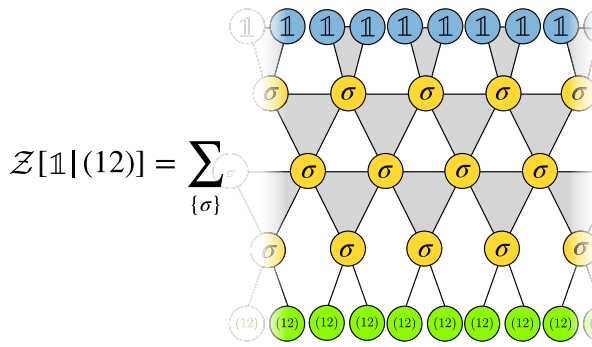


FIG. 28. Schematic: boundary conditions in the model for effective spins σ obtained for a 1D unitary circuit with or without measurements. Lines and triangles indicate interactions. The values of the boundary spins are fixed, and determined by the choice of contractions at the initial and final time in the multilayer circuit.

We do not need details of the lattice construction, but Fig. 27 shows an example for a 1 + 1D circuit geometry. For each unitary in the original circuit, we obtain a spin degree of freedom σ in the effective statistical mechanical model. We may write the partition function for these spins schematically as

$$\mathcal{Z}_N = \sum_{\{\sigma\}} W(\{\sigma\}). \quad (125)$$

The boundary conditions on the σ depend on the observable (Sec. IV B). The Boltzmann weight $W(\{\sigma\})$ is a product of local weights on each of the shaded triangles in Fig. 27: the form of the weight $J(\sigma_a, \sigma_b, \sigma_c)$ for the three spins $\sigma_a, \sigma_b, \sigma_c$ on a given triangle, specifying how they interact, may be found in Refs. [13,14] for a circuit with measurements and in Refs. [34,42] for the purely unitary case. Constructions for the random tensor network with random Gaussian tensors were discussed earlier in Ref. [35] and extended to take into account the replica trick in Ref. [36]. In all these cases the interaction terms are, loosely speaking, ferromagnetic, in that the Boltzmann weight is maximized when the σ configuration is uniform.

Physically, the spin σ should be thought of as a way to label a choice of *pairing* of the forward layers with the backward layers. Let the permutation $\sigma \in S_N$ map a given element $i \in \{1, \dots, N\}$ to $\sigma(i)$. Then σ stands for the pairing in which forward layer i is paired with backward layer $\sigma(i)$ and so on. For example the identity permutation, $\sigma = \mathbb{I}$, denotes the pairing of 1 with $\bar{1}$, of 2 with $\bar{2}$, and so on, i.e., in the pattern:

$$\sigma = \mathbb{I} : \quad \begin{array}{c} \frown \quad \frown \quad \frown \\ 1 \quad \bar{1} \quad 2 \quad \bar{2} \quad 3 \quad \bar{3} \end{array}. \quad (126)$$

We take $N = 3$ for this example, and we reorder the layers in comparison with Fig. 24 so the pairing can be drawn

without crossings. For the transposition, $\sigma = (12)$, layer 1 is paired with $\bar{2}$ and layer 2 with $\bar{1}$:

$$\sigma = (12) : \quad \begin{array}{c} \frown \quad \frown \quad \frown \\ 1 \quad \bar{2} \quad 2 \quad \bar{1} \quad 3 \quad \bar{3} \end{array}. \quad (127)$$

Since σ specifies a pairing of layers, we sometimes refer to it (and the continuum versions in the subsequent sections) as the “pairing field.” The physical interpretation of these pairings of layers is discussed in Sec. VID below. Heuristically, pairing Feynman histories in the discrete time evolution allows phase cancellation to be avoided [34,43], in the spirit of the diagonal approximation in periodic orbit theory [117].

That is, we may think of the multilayer circuit as a discrete path integral for N forward and N backward copies of the system. A Feynman trajectory is specified by a sequence of spin states in each of the copies. In a given layer, the corresponding product of matrix elements of local gates is the discrete analog of the exponentiated action for a continuum Feynman trajectory: e^{iS} or e^{-iS} depending on whether it is a forward or a backward layer. After averaging (or, in some cases, even without averaging [43]) this multilayer path integral may be dominated by configurations in which forward and backward layers form “pairs” with similar spin configurations, contributing opposite phases to the total weight. Such a pairing allows the effect of phase cancellation to be reduced. (See also Sec. VID.) The pattern of pairing will in general differ at different locations in space time, corresponding to space-time dependence of the pairing field σ . If the boundary conditions—say at the final time—involve pairwise index contractions of layers, as arise in the expressions for Rényi entropies (Sec. IV B), this will act as a boundary “magnetic field,” which selects out a particular value for the pairing field σ at the boundary (see for example Fig. 28).

G_N acts on σ via both left and right multiplications, for the two S_N factors, respectively, and via inversion for the

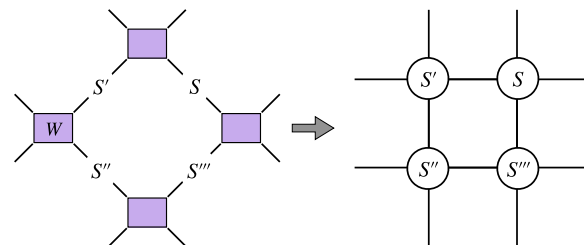


FIG. 29. Left: the simple random tensor network described around Eq. (147), with bond dimension 2, which reduces to a square-lattice Ising model with complex interaction constants (right). The Ising spin values on the right represent values of bond indices on the left.

\mathbb{Z}_2 generator, so that we have the symmetry transformations:

$$\sigma \rightarrow g_L \sigma g_R^{-1}, \quad \sigma \rightarrow \sigma^{-1}, \quad (128)$$

for permutations g_L and g_R , together with combinations of the above.

The effective spin interactions in Eq. (125) are local for simple choices of the random tensors or gates, but in general depend nontrivially on N , and may even be negative [120]. However, there is a relatively simple picture of the entangling phase as a phase where σ is ferromagnetically ordered, so that G_N is spontaneously broken, and of the disentangling phase as a disordered phase. Entanglement entropies may be expressed as free-energy costs for nonuniform boundary conditions [34–36] (see Sec. IV B). We appeal only to these facts and the symmetry structure above. Note that the simplest nontrivial case is $N = 2$: then there are only two possible pairings, \mathbb{I} and (12). Denoting these $+$ and $-$ leads to an effective Ising model [35,42]. In this case G_N reduces to a simple \mathbb{Z}_2 symmetry relating the two states.

Finally, we must specify the replica limits of interest. Loosely speaking, the required value of N [14,36] can be seen by counting powers of V . It is

$$N \rightarrow 1 \quad \text{for the MPT}, \quad (129)$$

$$N \rightarrow 0 \quad \text{for the RTN and FMPT}. \quad (130)$$

$N \rightarrow 0$ is what we typically have for systems with quenched randomness (Sec. VID). The additional power of V and V^* for the MPT comes from Born’s rule factor (recall that we label $V = V_{\mathbf{m}}$ by the sequence \mathbf{m} of measurement outcomes obtained in a given realization of the dynamics):

$$P(\mathbf{m}) = \langle \psi | V_{\mathbf{m}}^\dagger V_{\mathbf{m}} | \psi \rangle, \quad (131)$$

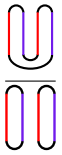
which must be included in every average for the MPT. We review this more carefully in Sec. IV B.

B. Boundary conditions in replica formalism

In order to review the replica formalism [13,14,36], let us express the operator entanglement S_2 of the nonunitary time-evolution operator V in a measurement or forced-measurement circuit. The latter case is precisely analogous to a random tensor network, except that for the case of time evolution there is a natural division of the external legs of the tensor network into those associated with the initial time and those associated with the final time. We focus in this subsection only on reviewing how the boundary conditions in the effective partition function arise formally (see

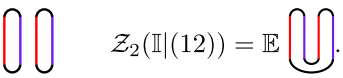
Sec. VID for more on how the “pairing field” arises in the bulk).

We define the operator entanglement in Sec. II. Recall that, if we view V formally as a tensor network *wave function* for $2N$ spins, then ρ_t is the unnormalized reduced density matrix associated with the final-time legs. Let us start with the case of the FMPT, where expectation values (denoted by $\mathbb{E}[\dots]$ or $[\dots]$), are simple averages over the unitaries and projections in V . The expectation value of the second Rényi entropy is

$$\overline{S}_2 = -\mathbb{E} \ln \frac{\text{tr} \rho_t^2}{(\text{tr} \rho_t)^2} = -\mathbb{E} \ln \frac{\text{Diagram}}{\text{Diagram}} \quad (132)$$


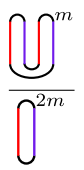
On the right we indicate the pattern of index contraction graphically. The vertical lines represent a stack of copies of V and V^* , like that in Fig. 24, but viewed from the side. For convenience, we order the four layers in the stack as follows: V^* , V , V^* , V (instead of grouping all of the V^* s together as we do in Fig. 24). The arcs at the top and bottom indicate the pattern of index contractions between layers. Index contractions are done separately for each of the physical sites.

Next let us define “partition functions” that are averages of the multilayer circuit with particular choices of boundary conditions. We use the notation $\mathcal{Z}_N(\sigma|\tau)$ for the average of the circuit with N layers of V and N layers of V^* , and with index contractions in the pairing pattern σ at the top and τ at the bottom. For example,

$$\mathcal{Z}_2(\mathbb{I}|\mathbb{I}) = \mathbb{E} \frac{\text{Diagram}}{\text{Diagram}} \quad \mathcal{Z}_2(\mathbb{I}|\text{(12)}) = \mathbb{E} \frac{\text{Diagram}}{\text{Diagram}}. \quad (133)$$


$\mathcal{Z}_N(\sigma|\tau)$ maps to a partition function for the pairing field with an effective “magnetic field” favoring pairing state σ at the final time (top) and τ at the initial time (bottom).

Equation (132) is not immediately written in terms of such partition functions, because of the logarithm and the fraction, but this can be dealt with using the replica trick [36]. Equation (132) is trivially equivalent to

$$\overline{S}_2 = -\frac{1}{m} \mathbb{E} \ln \frac{\text{Diagram}^m}{\text{Diagram}^{2m}} \quad (134)$$


for any $m > 0$, since the factors of m cancel. But one may check (by expanding in m in the numerator and denominator below) that in the limit $m \rightarrow 0$ the expectation value may be taken for the numerator and denominator

separately:

$$\bar{S}_2 = - \lim_{m \rightarrow 0} \frac{1}{m} \ln \frac{\mathbb{E} \left[\text{Diagram: two overlapping loops, one red, one blue, with a vertical line through them} \right]^m}{\mathbb{E} \left[\text{Diagram: two separate loops, one red, one blue} \right]^{2m}} \quad (135)$$

As usual, we treat m as a positive integer at intermediate stages of the calculation. The above then becomes

$$\bar{S}_2 = - \lim_{m \rightarrow 0} \frac{1}{m} \ln \frac{\mathcal{Z}_{2m}(\mathbb{I} | \tau_{2,m})}{\mathcal{Z}_{2m}(\mathbb{I} | \mathbb{I})}. \quad (136)$$

Here $\tau_{2,m}$ denotes a permutation in S_{2m} that is a product of m commuting two cycles [34]

$$\tau_{2,m} = (12)(34) \dots (2m-1, 2m). \quad (137)$$

Equation (136) may now be interpreted as the free-energy cost of imposing distinct boundary conditions for the pairing field σ (represented by the continuum field X in the sections below) at the initial and final times. If the free-energy cost for given boundary conditions σ and τ is [121] $\mathcal{F}_N(\sigma^{-1}\tau)$ then

$$\bar{S}_2 = \lim_{m \rightarrow 0} \frac{1}{m} \mathcal{F}_{2m}(\tau_{2,m}). \quad (138)$$

This generalizes directly to higher Rényi entropies. (The von Neumann entropy can either be obtained using an additional limit $n \rightarrow 1$, or by a slightly different construction with a single replica limit [36].) Note that the total number N of replicas (denoted $2m$ above) tends to zero as stated above for the FMPT and the RTN.

The simplest situation, discussed in the next subsection, is where the pairing field is well ordered across the entire sample. Then the free-energy cost \mathcal{F} is essentially the free-energy cost of inserting a single domain wall in this order [34–36]. See for example Fig. 30 in Sec. I.

In fact, in this situation (the strongly ordered regime) results from the unitary case suggest that in the replica limit can be dispensed with: we can map the entanglement to the free-energy cost of a single domain wall in an effective classical disordered system [34]. The most direct way to understand this is to avoid the replica trick entirely [43]. It is possible to make a formal mapping of the multilayer circuit in Eq. (132) (with $N = 2$) to an “Ising model” without any averaging. In general this model has complicated long-range interactions, so that it is not useful for discussing the critical point. But in the strongly ordered regime we expect (assuming the considerations for the unitary case in Ref. [43] carry over) that the interactions are effectively local after sufficient coarse graining. S_2 in a given realization can then be understood as a domain wall cost in a

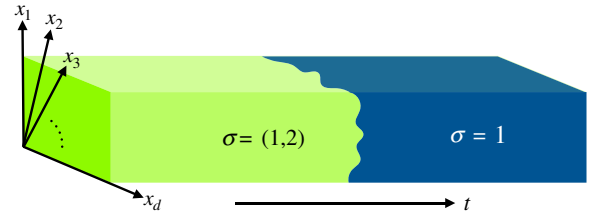


FIG. 30. At large times, a system in $d + 1$ space-time dimensions is quasi-one-dimensional. The long timescale τ in the entangled phase, with $\ln \tau \sim sN$, is due to the free-energy cost, which is approximately sN of a domain wall in the effective spin model (Sec. I). At early times, the replica trick can be avoided, giving a domain wall in a disordered Ising model. More generally we must use the replica spin model.

disordered Ising model. This is a route for justifying the use of an Ising model to discuss, for example, subleading corrections to the volume law in the entangled phase [24,27]. When the critical point is approached we must however return to the replica description above.

The application to the MPT is similar to the case of the FMPT. Graphically, the Born probability $P(\mathbf{m}) = \langle \psi | V_{\mathbf{m}}^\dagger V_{\mathbf{m}} | \psi \rangle$ for a sequence of measurement outcomes may be denoted by

$$P(\mathbf{m}) = \prod, \quad (139)$$

where the dots represent contraction with ψ or ψ^* as appropriate. Let us absorb a trivial constant into “ \mathbb{E} ” so that it denotes the average over the structure of the circuit together with the unweighted sum over \mathbf{m} :

$$\bar{S}_2 = - \mathbb{E} \left[\prod \ln \frac{\text{Diagram: two overlapping loops, one red, one blue, with a vertical line through them}}{\text{Diagram: two separate loops, one red, one blue}} \right]. \quad (140)$$

We may simplify the formulas slightly by averaging over the initial state, which yields

$$\bar{S}_2 = - \mathbb{E} \left[\prod \ln \frac{\text{Diagram: two overlapping loops, one red, one blue, with a vertical line through them}}{\text{Diagram: two separate loops, one red, one blue}} \right]. \quad (141)$$

The replica trick then allows us to write

$$\bar{S}_2 = - \lim_{m \rightarrow 0} \frac{1}{m} \ln \frac{\mathcal{Z}_{2m+1}(\mathbb{I} | \tau_{2,m})}{\mathcal{Z}_{2m+1}(\mathbb{I} | \mathbb{I})}. \quad (142)$$

Formally this is similar to Eq. (136), but the total number of replicas $N = 2m + 1$ is taken to 1 rather than 0 [13,14].

C. Permutations and coarse graining

We now focus on the critical properties. Let us first make a brief detour to consider coarse graining a lattice model of permutations, such as that shown schematically in Eq. (125), in an abstract sense, in order to understand one of the basic challenges. (This section is not an essential prerequisite for the following developments—the reader who wants to get to the concrete results may wish to skip it.)

We work throughout with a system in some finite number of dimensions $D = d + 1$ (the space-time dimension in the case of a circuit). Naively we might expect the limit of large d to match the all-to-all circuit (as in Sec. III E) but this is unclear (Sec. H).

Let us first imagine attempting a block-spin RG procedure in a naive way, by simply “averaging” the spins i within each D -dimensional local block:

$$\mu_{\text{block}} \propto \sum_{i \in \text{block}} \sigma_i. \quad (143)$$

What does this expression mean? At this point, each σ_i on the rhs is a formal group element in S_N . Their linear combination, μ_{block} , is no longer in S_N , since addition is not a group operation (only multiplication). Instead it is an element of the group algebra of S_N [122]. A general element of the group algebra is a linear combination of the elements g of the group with numerical coefficients M_g ,

$$\mu_{\text{block}} = \sum_{g \in S_N} M_g g, \quad (144)$$

where in the present case $M_g \in \mathbb{R}$. In other words, we can think of the coefficients M_g as forming a vector \mathbf{M} of length $N!$, which is the order of S_N . The coarse-grained spin above is equivalent to this vector.

However, μ_{block} , or equivalently the vector \mathbf{M} , is not a natural coarse-grained field in general. The reason for this is that \mathbf{M} does not form a single representation of the global symmetry G_N . Instead, the $N!$ -dimensional vector space splits into many distinct representations, in fact a number of representations that grows exponentially as N grows. Standard results for the group algebra imply that the representations of G_N that appear when we decompose μ_{block} are in one-to-one correspondence with the irreducible representations of S_N [122]. To extract a particular representation of G_N , we simply replace the formal group elements in Eq. (143) with their *matrix* representatives in the corresponding representation of S_N .

This means that our initial attempt to form a block spin has led us not to a single coarse-grained field, but to an indeterminate number (because N must be left free) of different coarse-grained fields, each in a different representation of the global symmetry group G_N .

In principle, we could try to write down a Lagrangian including all of these fields. However, since the number of these fields, and therefore the number of couplings, depends on N , this does not seem promising. Instead, it is natural to hope that only one or a small number of the fields become massless at the critical point, and the other fields do not need to be included in a continuum Lagrangian. This is the assumption we make, motivated by the more explicit picture in the following section.

This picture of splitting μ_{block} into separate fields gives an alternative view on the discussion of the percolation fixed point in Ref. [36]. The authors imagined starting with a lattice model with a much enlarged symmetry, $S_{N!}$ (not S_N or $S_N \times S_N$ or G_N). This much larger symmetry group is allowed to arbitrarily permute all the $N!$ values $\sigma \in S_N$ that the spin can take. Such a lattice model is simply a Potts model with $Q = N!$ states, for which the continuum theory is well known (becoming percolation when $Q \rightarrow 1$). The authors then considered deforming model in the direction of the physical model of interest [cf. Eq. (125)], which does *not* have $S_{N!}$ symmetry. They found that the lowest-order perturbation that could be added was quadratic in the Potts field, and so relevant. However, there was considerable freedom in the index structure of this perturbation, which could be formed from any class function of S_N .

From the present point of view, this perturbation is a sum of mass terms, with one independent mass for each of the infinite number of fields that appear when we decompose μ_{block} above into representations of S_N for arbitrary N .

D. Motivating a simple Landau theory

A familiar way to represent a permutation in S_N is as an $N \times N$ matrix $X_{a,b}$ of ones and zeros, with a single 1 in each row and in each column,

$$\sum_a X_{ab} = 1, \quad \sum_b X_{ab} = 1. \quad (145)$$

Under the global symmetries in Eq. (128), this matrix transforms as

$$X \rightarrow LXR^{-1}, \quad X \rightarrow X^T, \quad (146)$$

where L and R are permutation matrices representing g_L and g_R .

We might hope that we can build a Landau theory from such a matrix. In terms of the discussion in the previous section, this will correspond to the simplest choice of representations of G_N to include in the continuum theory (discussed below). In fact we can motivate such a Landau theory in a more direct way, without the need to go through the mappings discussed above involving permutations.

For this we appeal to the basic physical picture for why the pairings of layers arise in the multilayer circuit, which is to avoid phase cancellation. To make this explicit, let us

consider a particularly simple example of a tensor network V (which we can interpret formally as a nonunitary time evolution for qubits) with the geometry in Fig. 29(left). Label the bond index values by $S = \pm 1$ (these are the spins' σ_z values if V is interpreted as a time evolution). Take the local gates w , with bond indices S_1, S_2, S_3, S_4 , to have the simple form

$$w_{S_1, S_2, S_3, S_4} = \exp \left(\sum_{1 \leq i \leq 4} h_i S_i + \sum_{1 \leq i < j \leq 4} J_{ij} S_i S_j \right), \quad (147)$$

where each h is an independent, identically distributed *complex* Gaussian variable with mean zero, and equal variance $\Delta_h^2/4$ for its real and imaginary parts, and similarly for the J s, with variance $\Delta_J^2/2$. (These couplings are taken complex since tensors in a generic tensor network are complex.)

The tensor contraction defining V involves a sum over all the indices carried by the internal bonds in Fig. 29, i.e., over all Feynman trajectories, if we think of the vertical direction as time. This tensor contraction is an Ising partition function for the indices S_i on the bonds i . With the choices above, this Ising model lives on a rotated square lattice. We may write its partition function as

$$Z \equiv \sum_{\{S_i\}} \exp(i\mathcal{S}[\{S\}]), \quad (148)$$

where the exponentiated lattice “action” $e^{i\mathcal{S}}$ is just a product of terms of the form Eq. (147), so that $i\mathcal{S}[\{S\}]$ is an Ising Hamiltonian with random complex magnetic fields and random complex nearest-neighbor couplings. This is schematic as we have left the boundary conditions unspecified. (Fixed boundary conditions on the spins give a matrix element of V , for example; in practise we are interested in taking several layers of Z , which are coupled at their boundaries.)

Quantities of interest involve the replicated partition function (cf. Fig. 24). Up to boundary conditions, this is given by averaging $Z^N \times Z^{*N}$ over all of the random h and J parameters, as in the standard application of the replica trick to the Ising model with random bonds or random fields [82]. Introducing N replicas of the Ising spin for the forward layers, denoted S^a for $a = 1, \dots, N$, and N replicas for the backward layers denoted \bar{S}^b , the replicated partition function \mathcal{Z}_N has the form

$$\mathcal{Z}_N = \sum_{\{S_i^a\}, \{\bar{S}_i^b\}} \exp(-\mathcal{S}_{\text{eff}}) \quad (149)$$

(we do not include an i in the definition) with

$$\mathcal{S}_{\text{eff}} = -\Delta_J^2 \sum_{\langle ij \rangle} \sum_{ab} X_{ab}(i) X_{ab}(j) - \Delta_h^2 \sum_i \sum_{ab} X_{ab}(i), \quad (150)$$

where we define the “pairing field”

$$X_{ab}(i) = S_i^a \bar{S}_i^b. \quad (151)$$

This is similar to an Edwards-Anderson order parameter in an Ising spin glass. However, the usual Edwards-Anderson order parameter would be of the form $S^a S^b$ (as there would be no distinction between forward and backward layers) and the replica permutation symmetry would act on both a and b together. In the present case we have separate permutation symmetries for the row and column indices of X .

We defer an explicit discussion of coarse graining for this and other microscopic models to a separate publication [119]. Here we note only that the form of Eq. (150), with ferromagnetic interactions between the pairing field X for different sites, motivates writing a continuum Lagrangian for an $N \times N$ matrix, as discussed above.

In the present microscopic formulation, X is not a permutation matrix, but the action of symmetry is the same [Eq. (146)]. This is what we use, together with the assumption that the pattern of symmetry breaking in the entangled phase is the simplest one corresponding to pairing, i.e., to a choice of permutation.

Without loss of generality, let this permutation be the identity permutation (other cases are related by symmetry). Then the pattern of symmetry breaking is captured by an expectation value of the form

$$X_{ab} = f \delta_{ab} + c, \quad (152)$$

where f is the order parameter. Here c is a constant, which is generically nonzero even in the disordered phase, since (unlike f) it does not break any symmetry. This order breaks G_N down to $S_N \times \mathbb{Z}_2$, where the remaining permutation group is the subgroup of diagonal $S_N \times S_N$ transformations with $g_L = g_R$. (We briefly discuss more complex possibilities for symmetry breaking in Sec. J.)

The physical interpretation of X is simple: if in some region the spin configuration in the forward layer a is close to that in backward layer b , then the coarse-grained X_{ab} in this region will be large. Heuristically, we expect repulsive interactions between X_{ab} and X_{ac} for $b \neq c$: if the configuration in a is close to that in b , the phases from the a layer are already (partially) cancelled, so there is less gained by also pairing with c .

Let us briefly mention a caveat to the above discussion. A “random tensor network” is by definition a statistical

mechanics problem with very little required structure. Similarly the complex Ising model discussed above (which is an example of a random tensor network) is close to being the most general Ising model that one could write down for this lattice geometry [123]. On the other hand, the true measurement dynamics in the MPT does have some structure (for example, structure associated with causality), which is not present in a generic tensor network. In writing down the field theory in the next section we are assuming that the only aspect of the structure of the MPT that is important for the critical theory is the shift in the number of replicas from $N = 0$ to $N = 1$ that is induced by the Born probability. This assumption should certainly be examined further.

We note that the unitary limit, $r = 0$, is a case where additional structure due to unitarity certainly *is* important. There the appropriate effective “spin model” has hard constraints on the allowed spin configurations, which, for example, enforce causality [34,42,44] (these are relaxed when projection operators are included [13,14]). As a result, the unitary models do not possess invariance under $O(d + 1)$ rotations in space time, even in the scaling limit, and are not described by the field theories below, which do possess this symmetry. However, the unitary models do share some features with the ordered phases of these theories, such as a positive domain wall tension.

E. A field theory for the measurement transition

With this motivation, let us write the simplest Lagrangian for X_{ab} , which can represent a coarse graining either of a permutation matrix or of the composite field above. We see that this simplest Lagrangian passes a basic consistency check for the MPT. (In the next section we see that we need to extend it for the RTN and the FMPT.)

Let us make subtractions so that the row and column sums of the matrix give zero:

$$\sum_a \hat{X}_{ab} = 0, \quad \sum_b \hat{X}_{ab} = 0. \quad (153)$$

In the case where X is microscopically a permutation, this simply requires us to subtract a constant:

$$\hat{X}_{ab} = X_{ab} - \frac{1}{N}. \quad (154)$$

As a result of these linear constraints, which are preserved under coarse graining, \hat{X} has $(N - 1)^2$ independent components, and forms an irreducible representation of G_N . Below we omit the caret on \hat{X} .

Including terms in the potential only up to cubic order in X , and imposing G_N symmetry gives a relatively simple

Lagrangian. The theory we propose for the MPT is

$$\mathcal{L} = \sum_{ab} \left[\frac{1}{2} (\partial X_{ab})^2 + \frac{\mu}{2} X_{ab}^2 + g X_{ab}^3 \right]. \quad (155)$$

We include both time and space derivatives in the first term with the same coefficient, i.e., we set a nonuniversal speed to 1. This field theory has emergent Euclidean rotational invariance (not Lorentz invariance) in space time if this is not broken by boundary conditions. The components of the matrix X are not independent, because of the constraints in Eq. (153). Note that as a result, in contrast to the theory discussed in the next section, the only linear term $\sum_{ab} X_{ab}$ that would be allowed by symmetry is in fact zero. The replica limit $N \rightarrow 1$ is also implied. The renormalized squared mass vanishes at the critical point, $\mu^2 \propto (r - r_c)$.

Alternatively, we may write X in terms of an unconstrained $(N - 1) \times (N - 1)$ matrix field $\phi_{\alpha\beta}$ [124],

$$\mathcal{L} = \frac{1}{2} \sum_{\alpha\beta} [(\partial\phi_{\alpha\beta})^2 + \mu\phi_{\alpha\beta}^2] + g \sum D_{\alpha\beta\gamma}^{\mu\nu\lambda} \phi_{\alpha\mu} \phi_{\beta\nu} \phi_{\gamma\lambda}. \quad (156)$$

The tensor D is a tensor product of that appearing in the cubic term of the Potts model [80,81]:

$$D_{\alpha\beta\gamma}^{\mu\nu\lambda} = d_{\alpha\beta\gamma} d_{\mu\nu\lambda}, \quad d_{\alpha\beta\gamma} = \sum_{a=1}^N e_\alpha^a e_\beta^a e_\gamma^a. \quad (157)$$

The theory with the cubic term can only make sense for the replica limit—for $N > 2$ we have an unstable potential and for $N = 2$ the cubic term vanishes. This is also the case for the Landau-Ginsburg-Wilson-like theory for percolation, which we already discuss in Sec. III B. Like that theory, the upper critical space-time dimension of Eq. (155) is $D = 6$.

A basic consistency check on our picture is that this theory indeed sustains a stable ordered phase, with the simple pattern of symmetry breaking described in previous sections, when $\mu^2 < 0$. That is, the masses of fluctuations about the ordered state should remain positive in the replica limit $N \rightarrow 1$: otherwise some more complex pattern of symmetry breaking might be required [125–127]. To check this we put $X_{ab} = f(\delta_{ab} - 1/N) + W_{ab}$, where f is the magnitude of the order parameter, and W represents fluctuations (with $\sum_a W_{ab} = 0$, etc.). The saddle-point equation requires $f = \frac{-\mu^2}{3g} \frac{N}{N-2}$. The mass terms in the Lagrangian for W are then $\mathcal{L} = \frac{-\mu^2 N}{2(2-N)} (\sum_{ab} W_{ab}^2 - 2 \sum_a W_{aa}^2)$. We may check that the eigenvalues of the mass matrix appearing here are indeed positive when $\mu^2 < 0$ and $N \rightarrow 1$ (Appendix 1), so this consistency check is satisfied.

Now we consider another important consistency check.

F. Counting fields

Above we start with an $N \times N$ matrix X_{ab} transforming under G_N symmetry. For integer $N > 1$ we may split a general such matrix into four distinct fields, transforming under distinct representations of $S_N \times S_N$:

$$\begin{aligned} S &\equiv \sum_{ab} X_{ab}, \\ R_a &\equiv \sum_b X_{ab} - \frac{S}{N}, \\ C_b &\equiv \sum_a X_{ab} - \frac{S}{N}, \\ \hat{X}_{ab} &\equiv X_{ab} - \frac{C_a + R_b}{N} - \frac{S}{N^2}. \end{aligned} \quad (158)$$

The last of these, \hat{X} , is in the fundamental (standard) representation for both S_N factors. It lives in an irreducible representation of G_N of dimension $(N-1)^2$. R and C each transform under only one of the S_N factors. Since they are exchanged by the \mathbb{Z}_2 generator, together they form a single representation of G_N of dimension $2(N-1)$. S is a singlet.

For the MPT we construct a Landau theory that contained only the field \hat{X} . This is the obvious thing to do for various reasons (for example, if we think of \hat{X} microscopically as a permutation matrix, then R , C , and S are trivial constants). We conjecture that for the MPT \hat{X} is the only field that becomes massless at the critical point.

However, the group theory at $N \rightarrow 0$ [128] gives additional constraints, which strongly suggest that all of the representations in Eq. (158) become simultaneously massless at the critical point, so that we cannot throw away the representations R , C , and S . Therefore, we have to work with a general matrix X in which the row and column sums are not fixed to zero. The first indication of this is that the subtractions in Eqs. (154) and (158) diverge when $N \rightarrow 0$.

As with many other replica field theories, the partition functions that we are interested in become trivial—exactly equal to 1—in the replica limit, for certain choices of boundary conditions. An unusual feature of the circuit models with measurements or forced measurements is that this occurs at two values of N . When $N \rightarrow 0$ (FMPT) it occurs for the usual reason—because the partition function is the average of something raised to the power zero. When $N \rightarrow 1$ (MPT) the partition function is the sum of the probabilities of all the measurement outcomes—again giving 1 but for a different reason.

The fact that the microscopic partition function is equal to 1 implies constraints on the spectrum of operators in the continuum theory [118, 128–130]. Here a minimal heuristic point will be sufficient: there should not be any massless fields left when N is set equal to N_* , the desired number of replicas, otherwise we have a nontrivial free energy, contradicting $\mathcal{Z} = 1$.

The Lagrangian, Eq. (155), for the MPT satisfies this condition, since the field is in a representation of dimension $(N-1)^2$, which tends to zero when $N \rightarrow 1$. Therefore, it passes this basic consistency check.

At first we might assume that the same field theory can also be continued to $N = 0$ in order to describe the RTN and FMPT. However, this is not the case. Since $(N-1)^2$ is equal to one in this limit, rather than zero, this is not consistent.

However, the total multiplicity of all the representations in Eq. (158) is just N^2 (the number of components of the matrix), which *does* tend to zero in the replica limit $N \rightarrow 0$. This suggests that we should write a Lagrangian for a matrix X without imposing any condition on its row or column sums [131]. This is what we do next.

G. Field theory for random tensor network and FMPT

Let us denote the unconstrained real $N \times N$ matrix by Y , to distinguish it from the matrix X above, which obeyed linear constraints. Assuming only G_N symmetry and no constraints on Y , we argue below that the most relevant terms as $N \rightarrow 0$ are contained in

$$\mathcal{L} = \sum_{ab} \left[\frac{1}{2} (\partial Y_{ab})^2 + r Y_{ab} + g Y_{ab}^3 \right] + \frac{m_F^2}{2} \sum_{abcd} Y_{ab} F_{ab,cd} Y_{cd}, \quad (159)$$

where F is the tensor

$$F_{ab,cd} = \delta_{bd} + \delta_{ac}. \quad (160)$$

The parameter that drives this theory off criticality is r , the coefficient of the linear term (not to be confused with the measurement rate in previous sections, also denoted r). Since no constraint is imposed on Y , this linear term does not vanish (contrast Sec. E). The term $\sum_{ab} Y_{ab}^2$ is absent because its coupling can be set to zero by a shift $Y_{ab} \rightarrow Y_{ab} + C$ with a constant C , i.e., it is redundant [132]. Surprisingly, we find below that for this theory the upper critical dimensionality of space time is $D = 10$.

A peculiar feature of the $N \rightarrow 0$ limit of Eq. (159), which is shared with some other replica field theories such as the Landau-Ginsburg formulation of the random-field Ising model [82], is the presence of a quadratic coupling, which is *not* zero at the critical point and which cannot be removed. This is the term $m_F^2 Y.F.Y$.

If we instead study the above theory for a larger value of N , for example in the $N \rightarrow 1$ limit, then the effect of m_F^2 is simply to give a mass to certain representations in the decomposition of Y . The corresponding fields can therefore be eliminated at large scales and low momenta. Doing so returns us to the critical theory proposed in Sec. E for the MPT, with $\mu^2 \sim -r$. This is shown explicitly in Appendix 2. However, writing the propagator explicitly shows that

the limit $N \rightarrow 0$ that is of interest to us in this section does not commute with the limit of small momentum [82]. Therefore, we have to retain the YFY term explicitly.

Note that this term, which can be written

$$\sum_{abcd} Y_{ab} F_{ab,cd} Y_{cd} = \sum_a \left(\sum_b Y_{ab} \right)^2 + \sum_b \left(\sum_a Y_{ab} \right)^2,$$

includes contributions such as $Y_{12}Y_{13}$: this is consistent with the ‘‘repulsion’’ that is discussed heuristically towards the end of Sec. VID, between pairing patterns involving a given layer.

G_N symmetry allows many other terms at order Y^3 but we argue that in the $N \rightarrow 0$ limit they contribute only less relevant couplings. The dimensional analysis may be simplified using an approach [133–136] introduced by Cardy for the field theories of the random-field Ising model [137,138] and the branched polymer [139–141]. Since decomposition into representations of S_N fails in the $N \rightarrow 0$ limit, the next best thing is to exploit a decomposition into representations of an S_{N-1} subgroup acting on indices $2, \dots, N$. Here we must do this for both the row and column indices of Y .

We make a linear transformation to rewrite the field Y_{ab} as a field $y_{\alpha\beta}$ whose indices α and β take values in the set $\{+, -, 2, \dots, N\}$:

$$y_{\alpha\beta} = \frac{1}{2} \mathbf{v}^\alpha \cdot Y \cdot \mathbf{v}^\beta. \quad (161)$$

The index values $+$ and $-$ denote two distinct linear combinations that are invariant under S_{N-1} , while the values $2, \dots, N$ are permuted by S_{N-1} . The vectors \mathbf{v}^α are

$$\begin{aligned} \mathbf{v}^+ &= \frac{1}{2} \left(1, \frac{1}{N-1}, \dots, \frac{1}{N-1} \right), \\ \mathbf{v}^- &= \frac{1}{2} \left(1, \frac{-1}{N-1}, \dots, \frac{-1}{N-1} \right), \\ \mathbf{v}^i &= (0, \dots, 0, 1, 0, \dots, 0) - \frac{1}{N-1} (0, 1, \dots, 1), \end{aligned} \quad (162)$$

where in the last line the extra ‘‘1’’ is in the i th place. The $N-1$ vectors $\mathbf{v}^2, \dots, \mathbf{v}^N$ add up to zero, so span only an $N-2$ dimensional space. Below, indices i, j, k always run over $2, \dots, N$. Technical details are in Appendix 3.

After this rewriting, the terms in Eq. (159) up to quadratic order in Y become

$$\mathcal{L}_{\text{quadratic}} = \mathcal{L}^{(1)} + \sum_j \mathcal{L}_j^{(2)} + \sum_k \mathcal{L}_k^{(3)} + \sum_{jk} \mathcal{L}_{jk}^{(4)}, \quad (163)$$

with

$$\begin{aligned} \mathcal{L}^{(1)} &= (\partial y_{++}) (\partial y_{--}) + (\partial y_{+-}) (\partial y_{-+}) \\ &\quad + 2r y_{--} + 2m_F^2 y_{--} (y_{+-} + y_{-+}), \end{aligned} \quad (164)$$

$$\mathcal{L}_j^{(2)} = \frac{1}{2} (\partial y_{j+}) (\partial y_{j-}) + \frac{m_F^2}{2} y_{j-}^2, \quad (165)$$

$$\mathcal{L}_k^{(3)} = \frac{1}{2} (\partial y_{+k}) (\partial y_{-k}) + \frac{m_F^2}{2} y_{-k}^2, \quad (166)$$

$$\mathcal{L}^{(3)} = \frac{1}{4} (\partial y_{jk})^2. \quad (167)$$

Because of the linear constraints $\sum_j y_{j+} = 0$ etc., sectors (2) and (3) each contain $N-2$ copies of the same theory, and sector (4) contains $(N-2)^2$ copies of the same theory [142]. In the above rewriting, terms with couplings that vanish as $N \rightarrow 0$ are dropped [133].

Before writing the interaction terms, we use the quadratic terms to assign engineering dimensions to the various fields (see Appendix 3 for details). We assign dimensions $x_{\alpha\beta}$ such that all the quadratic terms in the Lagrangian Eq. (163) are marginal at the critical point $r = 0$. This gives

$$x_{\alpha\beta} = w_\alpha + w_\beta, \quad (168)$$

with (recall that in the case of a circuit $D = d + 1$ is the space-time dimension)

$$w_+ = \frac{D-6}{4}, \quad w_i = \frac{D-2}{4}, \quad w_- = \frac{D+2}{4}. \quad (169)$$

The RG eigenvalue of a cubic interaction term $y_{\alpha\beta} y_{\alpha'\beta'} y_{\alpha''\beta''}$ is then determined by the difference in the number of $+$ indices it contains and the number of $-$ indices it contains among α, \dots, β'' (Appendix 3).

However, the terms that can appear are constrained by the G_N symmetry (whose effects are less obvious in the new representation). We confirm in Appendix 3 that the cubic term $g \sum_{ab} Y_{ab}^3$ shown in Eq. (159) is strictly more relevant than the other symmetry-allowed cubic terms (at least for large enough D) and is of the form

$$\begin{aligned} g \sum Y_{ab}^3 &= \frac{g}{2} [6y_{++} (y_{++}y_{--} + 2y_{+-}y_{-+}) + 6y_{++} \\ &\quad \times (y_{-k}y_{+k} + y_{j-}y_{j+} + y_{jk}y_{jk}/4) \\ &\quad + 3(y_{+-}y_{j+}y_{j+} + y_{+-}y_{+k}y_{+k}) + 3y_{j+}y_{+k}y_{jk}] \\ &\quad + \text{less relevant terms.} \end{aligned} \quad (170)$$

The RG eigenvalue of g is $(10 - D)/2$, so the upper critical dimension for this theory is $D = 10$.

H. Consequences of the MPT field theory

We discuss some simple consequences of the putative field theory for the MPT, deferring a detailed analysis, and a discussion of the more complicated theory in the previous section, to another time. However, first we note an important caveat to the discussion.

Our initial hope is that the large- D limits of these field theories would give exact results both for the all-to-all circuits and for tree tensor networks. For example, this is what we find for the classical minimal cut toy model (because all-to-all percolation could be understood using the field theory for percolation in high dimensions, Sec. III E). But the class of tree tensor networks that we understand best, including those derived from the all-to-all FMPT circuit with Haar-random gates, seems *not* to be described by the field theory of Sec. G, simply because it is hard to imagine the exponential scaling of the order parameter in Eq. (52) being reproduced by a mean-field treatment of Eq. (159). Therefore, it seems unlikely that the all-to-all circuits studied in this paper are described by the $d \rightarrow \infty$ limit of the above field theories. We do not yet understand the reason for this difference.

It is not ruled out that our Lagrangians overlook some crucial structure, and that as a result they do not capture *any* models of measurement circuits or random tensor networks, even in finite dimensions. For present purposes we assume this pessimistic scenario does not hold, and that the two field theories in Secs. E and G do capture at least some class of models for the MPT and for the FMPT and RTN. We explore these issues further elsewhere.

The simpler of the two field theories is that in Sec. E for the MPT, involving a field X_{ab} with vanishing row and column sums. As a result of the cubic term, this theory has upper critical space-time dimension $d + 1 = 6$. Interestingly, the logic of Sec. III E for the percolation problem above five spatial dimensions applies in this case too, since it relies only on the engineering dimensions of the fields. We can therefore carry over the exact exponent values so that (neglecting physics on timescales shorter than $L = N^{1/d}$, see Sec. III E) the natural scaling variables in high dimensions are again

$$t/N^{1/5}, \quad N^{2/5}\delta r, \quad (171)$$

where $\delta r = r - r_c$ is the parameter driving the transition (and the number N of spins should not be confused with the replica number in the preceding sections).

Let us consider the operator entanglement in the ordered phase, still above the upper critical dimension. The plateau value of the operator entanglement, $S_2 \sim sN$, is proportional to the energy cost of a domain wall in X that spans the system in the spatial directions, as discussed in the following section. In high dimensions the scaling of s follows from dimensional analysis, giving $s \propto \mu^5 g^{-2}$, Eq. (155),

or in terms of the deviation δr from criticality,

$$s \sim |\delta r|^{5/2}, \quad (172)$$

which is the same exponent as for the classical problem in high dimensions. A similar scaling form will again apply, $\overline{S}_2 = H[t/N^{1/5}, N^{2/5}\delta r]$, but with a different scaling function H . The size of the order parameter X itself, which may be measured using appropriate correlation functions, grows linearly with the distance from the critical point, $X \sim |\delta r|$.

Again we have a characteristic timescale $\tau = N^{1/5}W(N^{2/5}\delta r)$, for an appropriate scaling function W . In the entangled phase this timescale grows exponentially in N (Sec. I), with

$$\ln \tau \sim sN \sim |\delta r|^{5/2}N \quad (173)$$

close to the transition.

Below $5 + 1$ dimensions the scaling is different, because the cubic term is no longer dangerously irrelevant. The appropriate scaling variables are as usual

$$t/L, \quad L^{1/\nu}\delta r, \quad (174)$$

where ν is the correlation length exponent for the field theory, Eq. (155). Exponents could be computed in a $6 - \epsilon$ expansion and will differ from percolation exponents (since the structure of the field theory is different, despite sharing the same upper critical dimension). In the ordered phase there is still an exponentially long timescale, with

$$\ln \tau \sim sN \sim |\delta r|^{\nu d}N \quad (175)$$

close to the transition (Sec. I).

I. Long timescale in the entangled phase

So far in this section we focus on the continuum description close to the transition. Here we discuss something simpler, namely the emergence of a timescale that (in the entangled phase) is exponentially large in the number of spins, and the contrasting short timescale in the disentangled phase. We may consider either a model in d spatial dimensions with $N = L^d$ spins, or the all-to-all model. The results in this section are independent of the conjectural field theories above, as they rely only on more basic features of the effective spin model (pairing field) descriptions.

The appearance of a long timescale may be understood in analogy to standard 1D or quasi-1D classical models. Here the 1D coordinate is time: see Fig. 30.

In the ordered phase the pairing field (either σ on the lattice or X in the field theory) has long-range order across a temporal slice and, after coarse graining sufficiently, we may think of it as a function only of time. There is then

a competition between the free-energy cost of imposing a domain wall at a particular time, which scales as sN with $s > 0$, and the entropy $\ln t$ associated with translating the domain wall in the time direction. At a timescale τ with

$$\ln \tau \propto sN \tag{176}$$

the translational entropy wins, and domain walls proliferate. Long-range order then no longer extends from the initial to the final time. By the identification of the entanglement with a free energy, this also means that the entanglement begins to decay exponentially with time.

Recently the exponentially long timescale in the entangled phase has been discussed from several points of view. References [12] and [56] consider a limit where the unitary evolution during a unit time can be treated as a $2^N \times 2^N$ Haar random unitary (see also Appendix E here for related considerations). Reference [27] has also given an analysis in terms of Ising domain walls that is similar to our considerations below.

The proportionality in Eq. (176) allows for an order-1 constant: however we expect that $N^{-1} \ln \tau$ vanishes in the same manner as s when the critical point is approached from the entangled side (for example, with the same power of the tuning parameter when this dependence is a power law).

At times sufficiently shorter than τ the operator entanglement entropy has a plateau at an extensive value. The plateau value is corrected by a negative subleading term whose magnitude grows logarithmically with time. In terms of the pairing field, the plateau regime is that where the number of domain walls is the minimal number allowed by the boundary conditions.

For S_2 , in the plateau regime, it is in fact sufficient to think about an Ising domain wall in a system with (Ising symmetric) disorder, for reason discussed towards the end of Sec. IV B. That is, we expect that the replica trick can be avoided in the strongly entangled regime. It is also possible to argue for the Ising picture using the replica treatment, by arguing that in this regime the replica theory is equivalent to the replica representation of a disordered Ising model [34]. In the replica treatment we have $m \rightarrow 0$ “elementary” domain walls, each associated with one of the transpositions in $\tau_{2,m}$ [Eq. (137)]. These may either bind together, forming a composite domain wall, or may separate for entropic reasons. See Ref. [34] for a discussion in the unitary case. However, this system of $m \rightarrow 0$ domain walls can really just be thought of as the replica description of a *single* Ising-like domain wall in a disordered environment. See Refs. [34,43] for details in the unitary case.

If we neglect quenched disorder, then we obtain

$$\overline{S_2} \sim sN - \ln t, \tag{177}$$

in the plateau region. The second term is the contribution from translational entropy, arising because the centre

of mass temporal coordinate t_{DW} of the domain wall can be located anywhere in $(0, t)$. The form in Eq. (177) was obtained in Ref. [12] in a limit of very dilute measurements, where the system can be viewed as completely scrambled by a random unitary between each measurement. Reference [27] gave a picture in terms of Ising domain walls equivalent to the one presented above. Here we also suggest how the effective Ising model can be justified (in an appropriate regime and at the level of universal properties) rather than being only a heuristic model. Our consideration also implies that we should take into account quenched disorder, as discussed below. (For another application of domain wall entropy in an effective 1D model to quantum chaos, see Refs. [117,143].)

As a check on the replica picture, we also consider a toy model for the entangled phase that involves multiplying large random matrices. A crude treatment in Appendix E (which neglects spatial structure, random fluctuations, and also the n dependence of the Rényi entropies) reduces to computing the singular values of a sub-block of a large Haar-random unitary. This treatment also yields Eq. (177), and shows that the plateau value sN determines the timescale for exponential decay of S_n in the regime of much later times, as expected from the above. An analysis of related random matrix models has recently been presented in Ref. [56].

Equation (177) is the simplest picture, neglecting quenched disorder. In reality there will be more complex crossovers. For example, in the all-to-all model there may be a regime of timescales where the subleading correction is not $\ln t$ but instead proportional to $\sqrt{N \ln t}$ as in the classical minimal cut problem (Sec. III D).

This is because the conditional free energy $F(t_{\text{DW}})$, given by fixing t_{DW} , will vary with t_{DW} due to randomness: $F(t_{\text{DW}}) = sN + \eta(t_{\text{DW}})$. In high enough dimensions, and therefore presumably also in the all-to-all model, the typical fluctuations $\eta(t_{\text{DW}})$ will be Gaussian with a scale \sqrt{N} . Although these fluctuations are much smaller than N , they are in principle much larger than 1. Therefore, at early enough times the free energy will be dominated by the optimal (most negative) value of $\eta(t_{\text{DW}})$, rather than by translational entropy (see endnote [144]). But at larger times, there may be a regime where $\ln t$ entropy again dominates, giving the functional form in Eq. (177). At still larger times multiple domain walls will proliferate (and the full replica treatment is required) and eventually S_2 decays exponentially in time.

The fact that only a single domain wall plays a role in the plateau regime means that there is an approximate factorization property for S_2 in a given realization of the circuit. If we divide V into two parts, $V^{(1)}$ corresponding to evolution from 0 to t' and $V^{(2)}$ from t' to t , then $e^{-S_2} \simeq e^{-S_2^{(1)}} + e^{-S_2^{(2)}}$. The first term includes configurations with $t_{\text{DW}} \in (0, t')$ and the second those with $t_{\text{DW}} \in$

(t', t) . (This is approximate not only because it neglects configurations with multiple domain walls but also because it does not correctly treat domain walls with t_{DW} close to t' .)

We now contrast the properties of the disentangled phase with those of the entangled phase. Let us take the limit $N \rightarrow \infty$ first, so that as usual we can define the operator entanglement per spin at a given time:

$$s_2(r, t) = \lim_{N \rightarrow \infty} \frac{\overline{S_2(r, t, N)}}{N} \quad (178)$$

(we write this equation for S_2 , but the choice of Rényi index $n \geq 1$ should not be crucial). In contrast to the quasi-1D limit discussed above, this is the free-energy cost, in an infinite slab of finite thickness, of imposing the domain-wall boundary conditions described in Sec. IV B. In the disentangled phase the free-energy cost per unit transverse area decays exponentially with the thickness of the slab, so that $s_2(r, t)$ decays exponentially to zero with time.

J. Variants and comments

In this subsection we discuss a few extensions of the field-theory approach we present, as well as some open questions.

The measurement problems and random tensor networks that we discuss so far have no internal global symmetries. One could also consider, say, measurement dynamics with an Ising symmetry [38,41]. The definition of the pairing field in Sec. VID allows such symmetries to be incorporated, and suggests that in many cases they will change the universality class of the entanglement transition.

For example, if the tensor network in Sec. VID has a \mathbb{Z}_2 Ising symmetry that changes the sign of S^a (and if we assume that the field whose mass vanishes at the transition is still $X_{ab} \sim S^a \overline{S}^b$) then odd powers of X are forbidden by symmetry in the continuum Lagrangian, which completely changes its structure in the limits of both $N \rightarrow 0$ and $N \rightarrow 1$.

This symmetry consideration highlights a feature of the discussion in Sec. VID, which is that the definition of X_{ab} involves choosing a local basis. In many cases this choice may not seem natural: for example, in many random models, the statistical invariance property emphasized in Sec. 1 ensures that any choice of local basis is equivalent to any other. (The exact mappings to models of permutations avoid having to choose a basis, but on the other hand it is less obvious how to coarse grain them.) An open question is whether this necessity of choosing a basis is just an aesthetic issue, or a fundamental one. Is it possible, for example, that the statistical invariance property imposes constraints on the continuum theory that we neglect to take into account?

Other restrictions on the unitaries, not related to conventional symmetries, can also change the symmetries of the replica theory. For example, if all the unitaries are real valued [145] then there is no distinction in the bulk between forward and backward layers. The symmetry group G_N is then enlarged to S_{2N} . In this case we can introduce a pairing field in a similar manner to Sec. VID, now with a replica symmetry action like that in standard disordered magnets and spin glasses. (The restriction to Clifford unitaries [7,12,15,17,22] is a more drastic change, which may require a different theoretical approach.)

The picture in Sec. VID relates random tensor networks (for which the limit $N \rightarrow 0$ is the appropriate one) to the language typically used to discuss spin glasses. This relation raises the question of whether other types of replica symmetry breaking, or other types of glass transition [127], are relevant to natural choices of circuit or tensor network. (Of course we could always engineer, say, a glassy phase if we specifically design a tensor network with this in mind.) For example, one could imagine a second transition taking place inside the entangled phase for some choices of tensor network. At the entanglement transition, the $2N$ layers form a collection of N pairs, breaking G_N symmetry down to $S_N \times \mathbb{Z}_2$. Can the residual S_N symmetry be broken in a subsequent transition? What are the entanglement properties of the resulting (presumably glassy) phase?

A statistical mechanics problem that provides a possible analogy for some of these phenomena is the directed polymer with random complex (or random sign) weights [146–151]. The replica formulation of this problem involves N copies of the polymer's partition function and N copies of its complex conjugate. Averaging over random phases forces the copies to form pairs in order to avoid phase cancellation [148], in analogy to the pairing phenomenon in the circuits. Further, the paired object—a bound state of polymers from different copies—may itself undergo phase transitions due to disorder. Perhaps this simpler problem can provide lessons for the circuit.

K. Free-fermion measurement dynamics

Models of free fermions subjected to stochastic dynamics [37,38,52–54,152–156] can also show a transition in $d > 1$ between two phases with differing amounts of entanglement [37]. However, instead of an area law and a volume law phase (for states in finite dimensions), we instead have an area law phase and a phase with a logarithmic violation of the area law [37,53,54].

We may also characterize the two phases by transmission of information between initial and final time, which gives a distinction that makes sense in any dimension or for the all-to-all setup. For concreteness we may consider the latter case. The model of Ref. [37], which used the language of Majorana fermions, has a simple field-theory description that is related to a model of classical

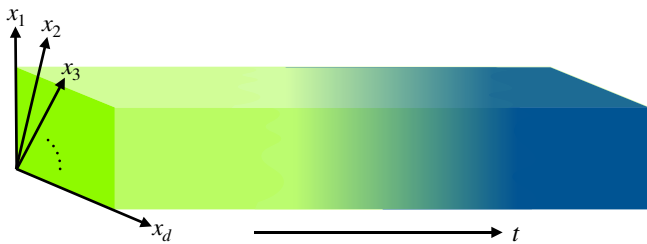


FIG. 31. The field-theory description of the Majorana measurement model of Ref. [37] has a continuous replicalike symmetry, allowing smooth domain walls that give a more rapid decay of S_n than in the interacting case where replica symmetry is discrete (Appendix B2). This exhibits a more general feature of free-fermion models.

loops (random walks) representing Majorana worldlines. The quasi-one-dimensional regime, which is relevant here has been studied in depth in Ref. [157], which also characterizes the statistical properties of random samples. Here we consider only some more basic average quantities.

The characteristic timescale for the operator entanglement to decay is of order N , where N is the number of lattice sites, rather than being exponentially large in N as we find in the interacting case. This is a generic feature of free-fermion models, as discussed below. Within the “more entangled” of the two phases, the scaling of the operator entanglement is

$$S_n \propto \begin{cases} K(r)N/t & t \ll K(r)N \\ \exp\{-ct/[K(r)N]\} & t \gg K(r)N. \end{cases} \quad (179)$$

Here K (the σ model stiffness) is an order-1 constant deep in the phase, and vanishes as $K(r) \sim (\delta r)^2$ upon approaching the transition at $r = r_c$ to the disentangled phase (c is a fixed order-1 constant). Note that the scaling in Eq. (179) is identical to the conductivity of a disordered N -channel wire, showing the crossover from Ohm’s law to localization on a timescale of order $K(r)N$ [157,158].

The reason for the reduced timescale in the entangled phase (of order N compared to the exponential timescale in interacting models) is that the appropriate replica field theory has continuous, rather than discrete, replica symmetry. In the ordered regime, a nonlinear σ model description may be used. Domain walls are smooth objects whose free-energy cost decreases with their thickness, which in the case of interest is the temporal duration t of the evolution: see Fig. 31.

For this reason, we anticipate that the scaling in Eq. (179) applies to more general free-fermion models with measurement. [The scaling of $K(r)$ close to r_c will depend on symmetries and dimensionality. The constant c may also depend on n in general.] General free-fermion models can be formulated using the replica trick, in close analogy to replica σ models for Anderson localization [159], leading to continuous replica symmetries. However,

in addition to the $N \rightarrow 0$ limit familiar from localization, the $N \rightarrow 1$ limit is now also of interest. We discuss this elsewhere [119].

The timescale of order N for free fermions agrees with the recent results of Ref. [56], which studied a model in which measurements of a single fermionic mode were alternated with Gaussian unitaries acting on the entire system. This model has even less locality structure than the all-to-all circuit. In this limit also, the authors found that $\mathcal{O}(N^2)$ measurements were required to forget the initial state: this corresponds to $t = \mathcal{O}(N)$ in our conventions.

VII. OUTLOOK

It remains an open question to what extent the properties of the MPT, in various settings, will turn out to be tractable (either analytically or numerically). In this paper, however, we show that exact results are possible in certain regimes. We close by summarizing the regimes we study, and some of the outstanding questions.

We begin our analysis by considering the “classical limit” of the MPT in the all-to-all setting. We show that a fairly complete picture is possible, including an analytical derivation of the critical point, critical exponents, and scaling forms for the entanglement.

Our results for quantum trees, including those obtained from a spin-1/2 all-to-all circuit, show that exact results are also possible even far from this classical limit. In this setting it is possible to demonstrate that an entanglement transition occurs at a definite nonzero measurement rate that is distinct from the classical value. (It may even be possible to obtain rigorous results on the phase diagram using the recursion relation approach.) The critical scaling on the tree is qualitatively different from a simple percolation picture.

We argue that the critical point on the tree is the same as the critical point of the FMPT in the all-to-all circuit (which is locally treelike). Since the location of the critical point in the circuit is difficult to check numerically, this equivalence has not yet been demonstrated clearly by our numerics. In the future we would like to have a clearer demonstration (or disproof) of this relationship between the tree and the all-to-all circuit. Our results based on the tree are also restricted to the FMPT; it would be interesting to understand to what extent they are relevant to the MPT.

The scaling on the tree raises several questions that we hope to return to elsewhere. First, it will be worthwhile to examine the relationship between the random recursion relation studied here and approaches to tree tensor networks based on replicas [68]. Second, we raise the question of whether there are multiple universality classes on the tree. This question remains to be settled, and could perhaps be addressed by generalizing our approach to a broader class of trees (with more general distributions of tensors or with larger bond dimension). Finally, it remains to be

understood how to reconcile the scaling that we find on the tree with field theory.

In our numerical study of the MPT we propose observables that have benefits over the state entanglement, in that they do not require one to specify a spatial sub-region. (Constructing such observables is crucial in the all-to-all setting, for which there is no meaningful distinction between area-law and volume-law scaling, but they are also useful in $1 + 1$ D, where significant finite-size effects make it important to avoid introducing length scales that are smaller than the system size.) We demonstrate numerically that there is a long timescale in the entangling phase over which some aspects of unitarity are retained; for example, two initially orthogonal states remain approximately orthogonal.

The optimal numerical protocol for studying critical properties in the all-to-all circuit remains to be settled. One complication is the lack of *a priori* knowledge of how the characteristic timescale scales with N when $r = r_c$. In the $1 + 1$ D problem, establishing that the dynamical exponent is equal to unity [6,15] allows one to reduce the number of independent variables in scaling collapses by fixing t/L to a constant. Our candidate field theory for the MPT suggests that in high dimensions the appropriate scaling variable is $t/N^{1/5}$, but it is unclear whether this theory applies to the all-to-all circuit.

The proximity of the classical critical point ($r_c^{\text{cl}} = 0.8$) to the quantum one (e.g., $r_c = 0.749$ for the FMPT with Haar-random gates) in the ensembles we study may also complicate the numerical analysis. For this reason it might be useful to study an all-to-all model (for example, involving weak measurements) in which the classical transition is eliminated entirely. It will also be interesting to relax the unitary invariance property of the gate distribution: the strong constraints imposed by this invariance are a surprising feature of our analysis of the quantum tree.

Finally, we address the replica approach to the MPT and to random tensor networks, both in the two phases and near the critical point, and we make concrete proposals for field theories for these problems. Determining the domain of applicability of these theories will be the subject of further work. A basic ingredient in Sec. VI is the construction of an “overlap” order parameter for the MPT and random tensor networks that brings these problems closer to the language we use for disordered magnets and spin glasses (in comparison with the more abstract language of permutation group elements used so far): this point of view may shed light on new possibilities for ordering.

ACKNOWLEDGMENTS

We thank Patrick Draper, John Chalker, Andrea De Luca, Tianci Zhou, Toby Swann, Michael Gullans, David Huse, Ehud Altman, Andreas Ludwig, Romain Vasseur,

Yi-Zhuang You, and David Kesler for useful discussions. A.N. is supported by a Royal Society University Research Fellowship. S.R. is supported by EPSRC Grant No. EP/S020527/1. J.R. acknowledges the funding of the Israeli Science Foundation under Grant No. 994/19 and the support of the Alon fellowship by the Israeli Council of Higher Education.

APPENDIX A: MORE ON CLASSICAL PROBLEM

1. Density of infinite cluster

Here we briefly derive Eq. (18), which describes the probability f_∞ that a given node in the interior of the classical graph is connected to an infinite number of other nodes in the limit of infinite N and T . In other words, f_∞ describes the density of the infinite cluster.

Consider the process of building a tree starting with an arbitrarily chosen node, as depicted in Fig. 4(c). The starting node has four possible edges, each of which may be severed by a measurement. If we denote by e_∞ the probability that following a given edge will lead to a subtree with an infinite number of nodes, then

$$f_\infty = 1 - (1 - e_\infty)^4. \quad (\text{A1})$$

The quantity $(1 - e_\infty)^4$ denotes the probability that *none* of the four edges connected to the starting node leads to an infinite number of other nodes.

Following a particular edge, one may next encounter either a measurement [with probability $p = r/(2 - r)$] or a node (with probability $1 - p$). The probability that this node is connected to an infinite number of other nodes at later generations is given by $1 - (1 - e_\infty)^3$. Thus we can write a self-consistency relation for e_∞ , given by

$$e_\infty = (1 - p)[1 - (1 - e_\infty)^3]. \quad (\text{A2})$$

Near the critical point, $p = 2/3 + \delta p$, where $\delta p = (25/18)\delta r$ and $\delta r = r - r_c \ll 1$. On the disconnected side of the transition, $e_\infty = 0$, while just on the connected side (small negative δr) $0 < e_\infty \ll 1$. Expanding Eq. (A2) for small δr gives $e_\infty \simeq -(25/6)\delta r$. A similar expansion of Eq. (A1) gives Eq. (18) of the main text.

2. Effective 1D field theory

Here we derive the mapping between the “layered Erdős-Rényi” percolation model and the one-dimensional field theory that is described in Sec. III B.

This is a bond percolation model with sites labeled (i, t) with $i = 1, \dots, N$ and $t = 1, \dots, T$. Generalizing slightly from the case in the text let a bond between sites (i, t) and (j, t) on the same timeslice be present with probability b/N , and a bond between sites (i, t) and $(j, t + 1)$ on the next slice be present with probability $b'/2N$. The average degree of a bulk node is $z = b + b'$, and from

considerations like those in Sec. A the critical case is $z = 1$.

Bond percolation can be simply mapped to the Potts model with $Q \rightarrow 1$ states (see Ref. [82] for a review). We introduce a Potts spin $\sigma(i, t) = 1, \dots, Q$ on each site (i, t) , and couplings for pairs of sites that are allowed to be connected by a bond. For each pair of spins that is allowed to be connected there is a term

$$[(1-p) + p\delta_{\sigma(i,t),\sigma(j,t)}] \quad (\text{A3})$$

in the Boltzmann weight, where p is the bond probability. The two terms correspond, in a diagrammatic expansion, to the presence and absence of the bond, respectively. Sites in the same percolation cluster have the same Potts spin state because of the Kronecker δ s on the bonds. Summing over spin states gives a factor of $Q^{\#\text{clusters}}$, which becomes 1 in the replica limit. Spin correlation functions can be used to diagnose connectivity. The probability that two sites (i, t) and (j, t') are in the same cluster is [82]

$$p_{\text{conn}}(i, t; j, t') = \lim_{Q \rightarrow 1} \frac{\langle \delta_{\sigma(i,t),\sigma(j,t')} - 1/Q \rangle}{1 - 1/Q}. \quad (\text{A4})$$

Below, the limit $Q \rightarrow 1$ will be left implicit.

Using the fact that the bond probabilities are of order $1/N \ll 1$, the partition function may be written

$$Z = \sum_{\{\sigma\}} \exp \left[\frac{b}{2N} \sum_{t=1}^T \sum_{ij} \delta_{\sigma(i,t),\sigma(j,t)} + \frac{b'}{2N} \sum_{t=1}^{T-1} \sum_{ij} \delta_{\sigma(i,t),\sigma(j,t+1)} \right]. \quad (\text{A5})$$

As is standard in the field-theory formulation of the Potts model [80,81], it is convenient to use a set of $(Q-1)$ -component vectors \mathbf{e}_σ , for $\sigma = 1, \dots, Q$ to represent the spin states, with the vectors satisfying

$$\mathbf{e}^\sigma \cdot \mathbf{e}^{\sigma'} = \delta_{\sigma,\sigma'} - Q^{-1}. \quad (\text{A6})$$

For $Q = 2$ we can take $e^{1,2} = \pm 1/\sqrt{2}$. For $Q = 3$ the three vectors point to the three corners of an equilateral triangle. For $Q = 4$ they point to the vertices of a regular tetrahedron, etc. Note that

$$\sum_{\sigma} \mathbf{e}^{\sigma} = 0, \quad \sum_{\sigma} e_{\mu}^{\sigma} e_{\nu}^{\sigma} = \delta_{\mu,\nu}, \quad (\text{A7})$$

as we see by considering $(\sum_{\sigma} e_{\mu}^{\sigma} e_{\nu}^{\sigma}) e_{\nu}^{\tau}$ and applying Eq. (A6). Writing $\mathbf{e}(i, t) = \mathbf{e}^{\sigma(i,t)}$, and denoting the sum of

the spins in a layer by

$$\mathbf{E}_t = \sum_i \mathbf{e}(i, t), \quad (\text{A8})$$

the partition function is (we drop an unimportant multiplicative constant)

$$\begin{aligned} Z &= \sum_{\{\sigma\}} \exp \left(\frac{b}{2N} \sum_{t=1}^T \mathbf{E}_t^2 + \frac{b'}{2N} \sum_{t=1}^{T-1} \mathbf{E}_t \cdot \mathbf{E}_{t+1} \right) \\ &= \sum_{\{\sigma\}} \exp \left[\frac{b-b'}{2N} \sum_{t=1}^T \mathbf{E}_t^2 + \frac{b'}{4N} \sum_{t=1}^{T-1} (\mathbf{E}_t + \mathbf{E}_{t+1})^2 + \frac{b'}{4N} (\mathbf{E}_1^2 + \mathbf{E}_T^2) \right]. \end{aligned} \quad (\text{A9})$$

We can use two sets of Hubbard-Stratonovich fields, one set located at half-integer times, denoted $\mathbf{f}_{t+1/2}$, to decouple the b' term, and one set located at integer times, denoted \mathbf{g}_t , to decouple the $b-b'$ term. Each has $Q-1$ components. Once the \mathbf{E} appear linearly in the exponent we can sum over the spins in a given timeslice t (the prime indicates that the sum is only over these spins) via

$$\sum'_{\{\sigma\}} \exp[\mathbf{E}_t \cdot \mathbf{y}] = \left(\sum_{\sigma=1}^Q e^{\mathbf{e}^{\sigma} \cdot \mathbf{y}} \right)^N = Q^N \exp[NV(\mathbf{y})], \quad (\text{A10})$$

which defines $V(\mathbf{y})$. Expanding in \mathbf{y} for small \mathbf{y} and using the identities mentioned above for the set of vectors $\{\mathbf{e}^{\sigma}\}$,

$$V(\mathbf{y}) = \frac{\mathbf{y}^2}{2Q} + \frac{d_{\mu\nu\lambda} y_{\mu} y_{\nu} y_{\lambda}}{6Q} + \mathcal{O}(y^4). \quad (\text{A11})$$

The tensor d is [80,81]

$$d_{\mu\nu\lambda} = \sum_{\sigma=1}^Q e_{\mu}^{\sigma} e_{\nu}^{\sigma} e_{\lambda}^{\sigma}. \quad (\text{A12})$$

After integrating out the spins,

$$Z = \int \mathcal{D}(f, g) e^{-\frac{1}{2}(\sum f^2 + \sum g^2) + \sum_{t=1}^T NV(\mathbf{y}_t)}, \quad (\text{A13})$$

where the final sum is over integer t . For $2 \leq t \leq T-1$,

$$\mathbf{y}_t = \sqrt{\frac{b'}{2N}} (f_{t-1/2} + f_{t+1/2}) + \sqrt{\frac{b-b'}{N}} g_t. \quad (\text{A14})$$

At the boundaries we have, e.g.,

$$\mathbf{y}_1 = \sqrt{\frac{b'}{2N}} f_{1+1/2} + \sqrt{\frac{b-b'/2}{N}} g_1. \quad (\text{A15})$$

For the present we neglect the boundary terms. The boundary condition on the field theory is important but below we fix it on physical grounds.

The negative power of N in \mathbf{y} will allow us to truncate the action at cubic order. Let us combine \mathbf{f} and \mathbf{g} into a field \mathbf{h} labeled by both integer and half-integer values, $\mathbf{h}_t = \mathbf{g}_t$, $\mathbf{h}_{t+1/2} = \mathbf{f}_{t+1/2}$. The lattice field theory is then (with $\tau, \tau' \in \mathbb{Z}/2$)

$$Z = \int \mathcal{D}\mathbf{h} \exp \left[-\frac{1}{2} \sum_{\tau, \tau'} h_\tau T_{\tau, \tau'} h_{\tau'} - \frac{1}{\sqrt{N}} A_3(\mathbf{h}) \right], \quad (\text{A16})$$

where A_3 contains the cubic terms. To avoid clutter, let us immediately set $Q = 1$ in the dispersion relation. The matrix T is then (the first row and column shown correspond to a half-odd-integer index value):

$$T = \begin{pmatrix} \dots & & & & & & & & \\ & 1 - b' & -\alpha/2 & -b'/2 & & & & & \\ & -\alpha/2 & 1 - b + b' & -\alpha/2 & 0 & & & & \\ & -b'/2 & -\alpha/2 & 1 - b' & -\alpha/2 & -b'/2 & & & \\ & & 0 & -\alpha/2 & 1 - b + b' & -\alpha/2 & & & \\ & & & & & & \dots & & \end{pmatrix} \quad (\text{A17})$$

Here we define

$$\alpha = \sqrt{2b'(b - b')}. \quad (\text{A18})$$

This becomes imaginary when $b' > b$ —which includes the line $b = 0$ on which we do simulations of this model—but this does not present a problem in the formal derivation below [160]. Let us write

$$b = \frac{z + \Delta}{2}, \quad b' = \frac{z - \Delta}{2}, \quad (\text{A19})$$

where z is the mean degree of a site, and the location of the critical point is $z = 1$ for any value of Δ . The dispersion relation has one “massive” mode, and one mode that becomes massless at the critical point $z = 1$, at frequency $\omega = 0$, with the eigenvalue of T being $(1 - z) + [(z - \Delta)/4]\omega^2 + \mathcal{O}(\omega^4)$. At $z = 1$ and $\omega = 0$ the eigenvector of this mode is $(g, f) \propto (\sqrt{\Delta}, \sqrt{1 - \Delta})$. For the low-frequency theory we make the coefficient of this mode a slowly varying field, $\phi(t)$. Let us write

$$\delta z = z - 1 \quad (\text{A20})$$

for the parameter that vanishes at the phase transition. Let us drop the small parameter δz except in the mass term, where it is the leading factor:

$$Z = \int \mathcal{D}\phi \exp \left\{ -\int dt \left[\frac{1 - \Delta}{8} (\partial_t \phi)^2 - \frac{\delta z}{2} \phi^2 \right] - \frac{A_3}{\sqrt{N}} \right\}. \quad (\text{A21})$$

Thanks to the small prefactor $1/\sqrt{N}$ of the cubic term, we may take the continuum limit in a controlled manner. The cubic term is negligible for frequencies ω of order 1 due to

the small prefactor, but important at parametrically small frequencies (since it is RG relevant). Since only small frequencies are important we can simply insert the form of the low-lying mode at $k = 0$ into the cubic term without any need to explicitly integrate out high-frequency modes. The final result is

$$Z = \int \mathcal{D}\mathbf{h} \exp \left(-\int dt \mathcal{L} \right), \quad (\text{A22})$$

with the “Lagrangian” (again a factor of Q has been set to 1 in the denominator of the final term)

$$\mathcal{L} = \frac{1 - \Delta}{8} (\partial_t \phi)^2 - \frac{\delta z}{2} \phi^2 - \frac{d_{\mu\nu\lambda}}{6\sqrt{N}} \phi_\mu \phi_\nu \phi_\lambda. \quad (\text{A23})$$

Above, all the $(Q - 1)$ components of the field ϕ are independent. We can write a more explicit form at the cost of using Q fields that obey a linear constraint (summing to zero). For notational convenience we write them as the components of a $Q \times Q$ diagonal matrix with components

$$\Phi_{\sigma, \sigma} = \phi \cdot \mathbf{e}^\sigma. \quad (\text{A24})$$

The constraint is tracelessness

$$\text{tr} \Phi = 0. \quad (\text{A25})$$

Equation (A23) becomes

$$\mathcal{L} = \frac{1 - \Delta}{8} \text{tr}(\partial_t \Phi)^2 - \frac{\delta z}{2} \text{tr} \Phi^2 - \frac{1}{6\sqrt{N}} \text{tr} \Phi^3. \quad (\text{A26})$$

This is the result given in the main text for the special case $b = 0$ (i.e., $\Delta = -z \simeq -1$ close to the critical point).

Close to the critical point, the connectedness correlation function for sites at distinct times is

$$p_{\text{conn}}(i, t; j, t') = \frac{\langle \vec{\phi}(t) \cdot \vec{\phi}(t') \rangle}{N(1 - 1/Q)} \rightarrow \frac{\langle \phi_1(t) \phi_1(t') \rangle}{N}, \quad (\text{A27})$$

where ϕ_1 is an arbitrarily chosen component. We can also write this as

$$p_{\text{conn}}(i, t; j, t') = \frac{\langle \text{tr} \Phi(t) \Phi(t') \rangle}{N(Q - 1)}. \quad (\text{A28})$$

In Appendix A 4 we present results for the connectedness correlation functions of boundary points. Since the boundary conditions on the Potts spins are free, this corresponds to the ‘‘ordinary’’ surface transition (discussed for percolation in Refs. [161–163]) where the boundary spin operator is $\propto \partial_t \Phi$ in the continuum theory rather than Φ as in the previous equation [82]. This gives the scaling forms in Appendix A 4.

Finally let us consider the percolation probability P_{perc} . This can be used to define a characteristic timescale $t_*(r, N)$ for the classical problem, and it is much simpler to formulate in field theory than the minimal cut cost. As discussed in Sec. III E, all this carries over to finite spatial dimensions $d > 5$ by setting $N = L^d$. P_{perc} is equal to $1 - e^{-\Delta F}$, where ΔF is the free-energy cost of imposing twisted boundary conditions [164] on the Potts spins [84].

The scaling form is

$$P_{\text{perc}} = F(t/N^{1/5}, \delta r t^2), \quad (\text{A29})$$

where t is now the total time, and we use the notation r for the parameter driving the transition to match the circuit. Let us consider a few different regimes.

By a rescaling of the field and the time coordinate we can choose to write the action in the form (suppressing order-1 constants)

$$S_{\text{eff}} = \frac{N}{f^5} \text{tr} \int_0^1 du [(\partial_u \tilde{\Phi})^2 + (\delta r t^2) \tilde{\Phi}^2 + \tilde{\Phi}^3]. \quad (\text{A30})$$

This rewriting suggests that if we take the limit of large N and t (and small δr) in such a way that the scaling variable $\delta r t^2$ is fixed while N/t^5 becomes large, ΔF is given by a saddle-point action (we do not try to make this precise in this replica theory),

$$\ln(1 - P_{\text{perc}}) \sim -\frac{N}{f^5} A(\delta r t^2), \quad (\text{A31})$$

In particular, at the critical point

$$\ln(1 - P_{\text{perc}}) \propto -\frac{N}{f^5} \quad \text{for } r = r_c \text{ and } t \ll N^{1/5}. \quad (\text{A32})$$

Note that for a system with finite $d > 5$, unlike the case $d < 5$, P_{perc} is parametrically close to 1 at the critical point

of a system with $t \sim L$, i.e., with $t \sim N^{1/d}$. This is because, for percolation above the upper critical dimension, there are many percolating clusters in a large hypercubic sample at p_c [165–170].

Next, let us take N large with $\delta r < 0$ small but fixed, in order to examine the exponential growth of the timescale with N inside the percolating phase. By an alternative rescaling of the field,

$$S_{\text{eff}} = N \text{tr} \int_0^t dt' [(\partial_{t'} \hat{\Phi})^2 + (\delta r t^2) \hat{\Phi}^2 + \hat{\Phi}^3]. \quad (\text{A33})$$

Assuming again that we can make an analogy with saddle-point solutions in more conventional theories with discrete symmetry, we anticipate a localized domain wall or ‘‘instanton’’ solution interpolating between the two boundary condition values of the spin, with a classical action $N \times c(r)$. The scaling form will require $c(r) \sim (\delta r)^{5/2}$ close to the critical point. At sufficiently early times there is at most one such instanton, which can be placed at any time in between 0 and t :

$$1 - P_{\text{perc}} \sim t |\delta r|^{1/2} \exp[-\text{const}(\delta r)^{5/2} N]. \quad (\text{A34})$$

The δr dependence of the prefactor has been fixed by requiring consistency with the scaling form [171]. Therefore, the plateau at $P_{\text{perc}} \simeq 1$ lasts for an exponentially long time

$$t_* \sim \frac{1}{\sqrt{|\delta r|}} \exp[\text{const}(\delta r)^{5/2} N]. \quad (\text{A35})$$

The interpretation is just that the probability of having a disconnection event at a given time is $p_{\text{break}} \sim |\delta r|^{1/2} e^{-\text{const}(\delta r)^{5/2}}$. Since the probabilities of such events are independent, at long times we have exponential decay of P_{perc} , with a timescale also given by Eq. (A35).

3. Criticality in layered Erdős-Rényi graphs

As introduced in Sec. III B, the layered Erdős-Rényi model is a simplification of the classical random graph depicted in Fig. 4(b), in which a large number N of nodes are arranged in discrete layers with time index t . Each node may be connected only to nodes in adjacent time layers; there are no connections between nodes within the same time layer. Edges between time layers t and $t + 1$ are randomly chosen, such that a total number of edges cN are created between adjacent layers. The connectivity c is the major parameter of the model (c is equal to $b'/2$, in the notation of Sec. III B), and plays a similar role as the complement of the measurement rate, $1 - r$. The critical value of c is $c_{\text{crit}} = 1/2$, since a given node at time t has connections to both $t - 1$ and $t + 1$ and its expected number of total connections is $2c$.

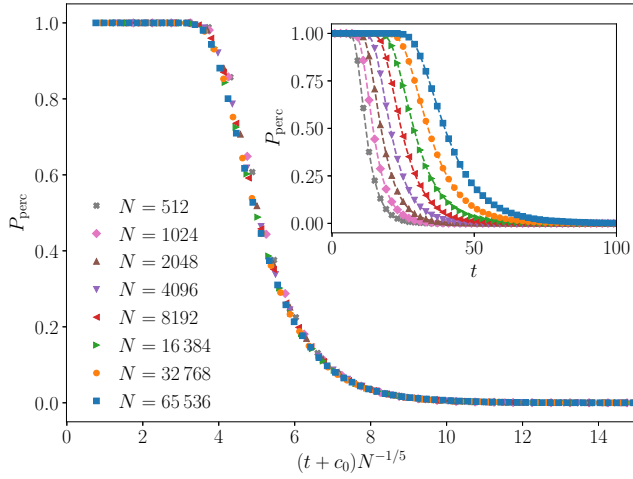


FIG. 32. The percolation probability P_{perc} for the layered Erdős-Rényi graph at the critical connectivity $b'_{\text{crit}} = 1$ is plotted as a function of time. The inset shows the raw data for different system sizes N , and the main figure shows the same data plotted as a function of $(t + c_0)/N^{1/5}$, where $c_0 = 6$ is a constant. Compare Fig. 5 of the main text.

Since the layered Erdős-Rényi description is the basis for the theoretical derivation of the scaling forms in Sec. III B, we numerically simulate the layered Erdős-Rényi graph to verify these scaling forms and ensure that they yield the same behavior as the data presented in Figs. 5 and 6 for the full classical graph.

Figure 32 shows the percolation probability at the critical point, plotted as a function of time for different system sizes N . A good scaling is observed as a function of the variable $t/N^{1/5}$, as suggested by Eq. (19). At a fixed value of c and N , the percolation probability P_{perc} is observed to decay exponentially with time. As shown in Fig. 33, near the critical point the scaled exponential decay time $\tau(c, N)/N^{1/5}$ is a function only of the variable $(c - c_{\text{crit}})N^{2/5}$. This is consistent with the scaling forms in Eq. (19).

4. Two-point correlation functions

In Sec. III C we show that the probability of percolation in the classical graph can be described in terms of the scaling variables $t/N^{1/5}$ and $(r - r_c)N^{2/5}$ [Eq. (19)]. Various correlation functions can also be understood in terms of these same scaling variables. The simplest correlation function, which we denote by C , is the probability that two distinct nodes in the graph belong to the same cluster. The nodes may be on a temporal boundary (either the same boundary or different ones) or in the bulk of the graph.

In the Potts language, the correlator is the spin two-point function. The bulk and boundary operators have different scaling dimensions, with the former scaling like $N^{-2/5}$ (or equivalently like t^{-2}) and the latter like $N^{-3/5}$ (or t^{-3}); see Sec. III E and Appendix A 2.

As a result of this scaling, the probability C_{oppo} that two nodes on opposite temporal boundaries are connected by a cluster has the form

$$C_{\text{oppo}}(T, N) = \frac{1}{N^{6/5}} F_{\text{oppo}}(T/N^{1/5}, \delta r N^{2/5}). \quad (\text{A36})$$

In Fig. 34 we test this scaling for case $r = r_c$, in order to confirm the theoretical value for the operator's scaling dimension. This data is for the simplified multilayer Erdős-Rényi model described in Sec. III B.

The probability C_{same} for two sites on the *same* temporal boundary to be connected has, in addition to the scaling term, a noncritical contribution of order $1/N$, which is in fact dominant at r_c . This $1/N$ factor is on the order of the probability for the two sites to be connected by a “microscopic” path (for example, by a single bond). For simplicity, consider the limit $T \rightarrow \infty$, when only one of the arguments of the scaling function remains:

$$C_{\text{same}}(N) = \frac{1}{N^{6/5}} F_{\text{same}}(\delta r N^{2/5}) + \frac{A(r)}{N}. \quad (\text{A37})$$

The noncritical term may be eliminated by a subtraction: $\tilde{C}_{\text{same}}(N) = C_{\text{same}}(N) - 2^{-1} C_{\text{same}}(N/2)$. Figure 35 demonstrates a reasonable scaling collapse for this quantity. This plot constitutes a second check that $\delta r N^{2/5}$ is the appropriate off-critical scaling variable.

If $r \lesssim r_c$ is fixed and $N \rightarrow \infty$, Eq. (A37) shows that $C_{\text{same}}(N)$ scales like $(r_c - r)^3$; this is the square of the *surface* order parameter (the probability that a boundary site lies in the infinite cluster), which is parametrically

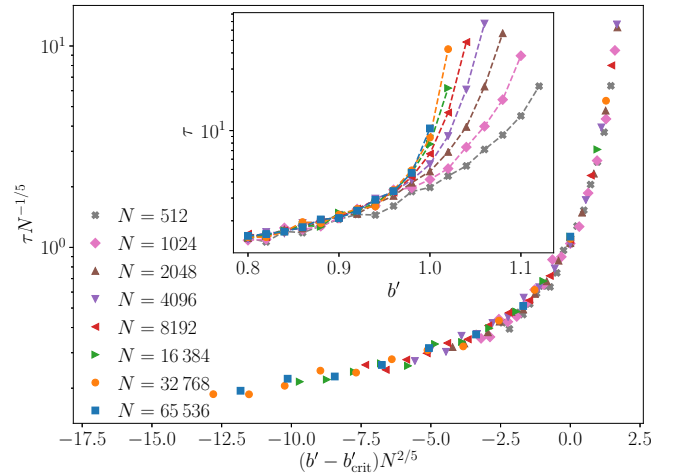


FIG. 33. The timescale τ for exponential decay of the percolation probability is plotted as a function of the connectivity b' and system size N for the layered Erdős-Rényi graph (inset). This timescale exhibits critical scaling when plotted as a function of the scaled variables $\tau/N^{1/5}$ and $(b' - b'_{\text{crit}})N^{2/5}$, as suggested by Eq. (19) with critical connectivity $b'_{\text{crit}} = 1$. Compare Fig. 6 of the main text.

smaller than the bulk order parameter [which scales as in Eq. (18)] when r_c is small. In the language of surface critical phenomena, this is the “ordinary” transition [82,161–163].

As an aside, let us make a distinction between the correlation function C_{same} above and the mutual information between spins in the final state. C_{same} indicates whether in the final state two spins lie in the same connected tensor network. However, being in the same connected component does not imply that spins’ mutual information, which can be detected with appropriate physical two-point functions, is large [6]. The zeroth Rényi mutual information I_0 is given by a different classical correlation function to the one above, which in the finite-dimensional problem behaves as a power law at r_c [6]. A related observable is the distribution of entanglement entropy S_0 for a single spin [17]. These observables again map to boundary correlation functions of the Potts spins [172], but in a Potts system with a magnetized boundary condition, rather than with the free boundary conditions used above for C_{same} . (In the classification of surface criticality this is the “extraordinary” transition [173,174].) The critical contribution to these quantities at r_c is smaller than a trivial analytic contribution similar to that mentioned above, we have not been able to see it numerically.

5. Extrapolating min-cut tension to $N \rightarrow \infty$

As mentioned in Sec. III D, the behavior of the classical minimum cut value $S_0(t, N)$ within the percolating phase,

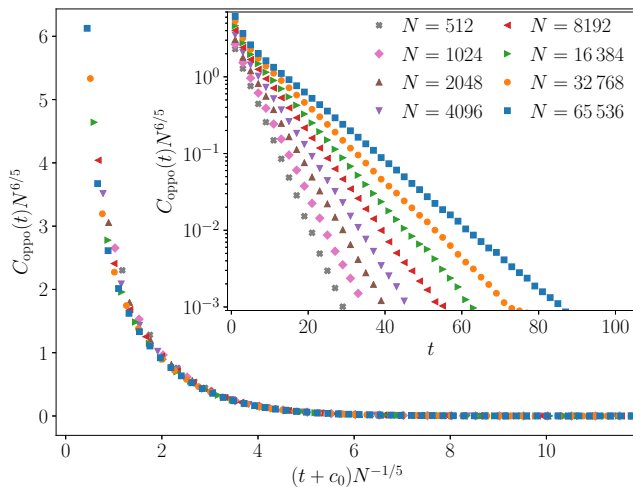


FIG. 34. The correlation function $C_{\text{oppo}}(t)$, which denotes the probability of two randomly chosen nodes on opposite temporal boundaries being connected to the same cluster after a total time evolution t . Data here corresponds to the layered Erdős-Rényi model at the critical point, $c = 1/2$. The inset shows $C_{\text{oppo}}(t)N^{6/5}$ as a function of t for different system sizes N , while the main figure shows that this same data scales onto a single curve when plotted as a function of $(t + c_0)N^{-1/5}$. Here $c_0 \approx 3.1$ is a constant.

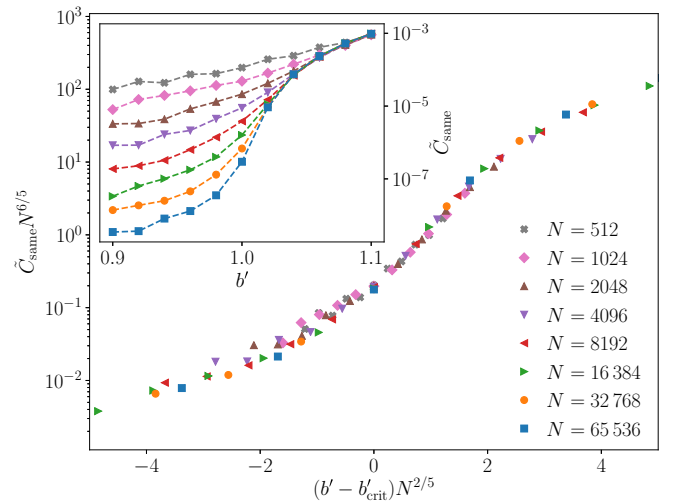


FIG. 35. The critical contribution \tilde{C}_{same} to the probability for two sites on the same temporal boundary to be connected, in the limit of long simulation time T . Data corresponds to the layered Erdős-Rényi model, and is plotted against b' , twice the expected number of connections between a node at time t and all other nodes at time $t + 1$. The inset shows the raw data for different system size N , while the main figure shows the data plotted in terms of scaled variables.

$r < r_c$, has the functional form

$$\frac{S_0(t, N, r)}{N} \simeq s(r) - \frac{d(t, r)}{\sqrt{N}}, \quad (\text{A38})$$

at large N , and for all times t that are larger than an initial transient but short enough to satisfy $\ln t \ll \ln \tau$. Here, $d(t, r)$ is a constant for fixed t and r and τ denotes the decay time of the percolation probability; τ grows exponentially with N (see Fig. 6). An example of this scaling is shown in Fig. 36. One can estimate the value of the minimum cut per spin, $s(r)$, by extrapolating this relation to $1/\sqrt{N} = 0$. As illustrated in Fig. 36, this extrapolated value is relatively insensitive to the time t , so long as t is larger than a short-time transient ($t \gtrsim 5$ is sufficient for the r values plotted here and in Fig. 7) and shorter than τ .

The data in Fig. 7 corresponds to $t = 10$, and comes from an extrapolation using system sizes $N = 100, 200, 400, 600, \text{ and } 800$. As shown in Fig. 36, performing an extrapolation at $t = 20$ yields essentially identical results. The extrapolation procedure becomes numerically difficult at very small $r_c - r$, since the decay time τ becomes short for all but very large system sizes. The extrapolated data in Fig. 7 is therefore limited to $r_c - r \geq 0.02$.

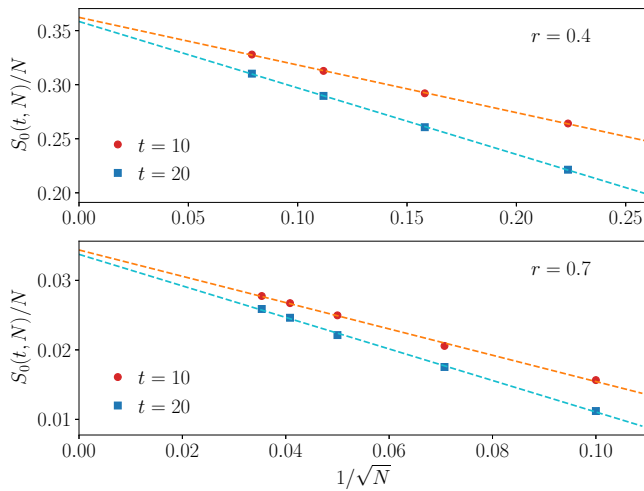


FIG. 36. An example of the extrapolation of $S(t, N)/N$ to the limit of $N \rightarrow \infty$ in order to estimate the classical minimal cut per spin, $s(r)$. Points correspond to S_0/N for different simulated system sizes at two different times t , and the dashed lines are linear best fit lines to S_0/N as a function of $1/\sqrt{N}$. The y intercept of these lines gives the estimated value of $s(r)$, plotted in Fig. 7. The extrapolated values of $s(r)$ for two values of t agree to within 3%. The top plot corresponds to $r = 0.4$, and the bottom plot corresponds to $r = 0.7$.

APPENDIX B: MAJORANAS WITH PAIRWISE MEASUREMENT

Reference [37] describes measurement-only dynamics for an even number of Majorana modes γ_i for $i = 1, \dots, N \gg 1$. Each measurement is of a fermion parity $i\gamma_i\gamma_j$ for some i and some j . (If desired we can impose a bipartite structure, so that measurements are allowed only for $i \in A$ and $j \in B$, where A and B are say two sublattices of a bipartite lattice.) In finite dimensions, such a model allows a phase transition between an area-law phase and a phase with a logarithmic violation of the area law. We briefly note the fate of this transition in the all-to-all setting.

We first select a single preferred grouping of the Majoranas: for i odd, i is grouped with $i + 1$. We can think of the two Majoranas within a group as forming a single complex fermion operator. With probability r , a measurement is an *intragroup* measurement, and with probability $1 - r$ it involves two Majoranas chosen uniformly at random, so that it is an *intergroup* measurement with probability 1 in the limit $N \rightarrow \infty$.

The trivial phase arises at large r , when intragroup measurements predominate, and the nontrivial phase at small r when intergroup measurements predominate. The dynamics can be thought of as the dynamics of an evolving *pairing* between Majorana modes [37]. In space time it maps to a loop model [175,176]. The loops represent Majorana worldlines, and the operator entanglement is proportional

to the number of worldlines connecting the initial to the final time. The relevant case for the all-to-all model is a quasi-1D version of such a loop model [157].

The phase transition is described by the RP^{n-1} σ model in the limit $n \rightarrow 1$ (or CP^{n-1} if we impose a bipartite structure). The relation between the all-to-all case and the finite-dimensional case is similar to that in Sec. III E. For the critical point it is natural to use a soft-spin formulation of the RP^{n-1} or CP^{n-1} model, with a cubic term and upper critical dimension 6 [176,177], so that by the logic of Sec. III E we expect the scaling variables in Eq. (19) to apply. However, the associated scaling functions, and some of the other exponents, will be very different. The basic difference from the field theory discussed in Sec. III E is that the RP^{n-1} and CP^{n-1} models have continuous replicalike symmetries [respectively, $SU(n)$ and $SO(n)$], unlike the discrete symmetries in both the field theory for percolation (Sec. III E) and the replica descriptions of the generic quantum problem (Sec. VI). This difference leads to much smaller entanglement in the nontrivial phase.

The entanglement in the present model is related to the free-energy cost of twisted boundary conditions in the σ model. The possibility of continuous twisting of the order parameter makes this smaller than in a system with discrete symmetry. In finite dimensions, the nontrivial phase has only a logarithmic violation of the area law, rather than volume-law entanglement [37].

In the all-to-all setting we again characterize the phases by the operator entanglement between initial and final times. In the nontrivial phase the RP^{n-1} field is “ordered” (on timescales much shorter than τ below) and so we use a nonlinear σ model formulation for an $n \times n$ matrix Q_{ab} that parameterizes RP^{n-1} . This has the form $S = K \int dt \text{tr}(\partial_t Q)^2$, with a stiffness that scales as

$$K \propto (\delta r)^2 N \quad (\text{B1})$$

close to the transition on the nontrivial side. A rescaling of t shows that this leads to a characteristic timescale of order

$$\tau \sim (\delta r)^2 N. \quad (\text{B2})$$

For $t \gg \tau$, the simple one-dimensional field theory has exponentially decaying correlations, and the operator entanglement (computed from the cost of imposing twisted boundary conditions between the initial and final time [178]) decays exponentially. For $t \ll \tau$, the operator entanglement is extensive in N and scales as $S \sim \delta r^2 N/t$.

Note the large difference between the timescale in Eq. (B2), which is linear in N , and the timescale in the entangled phase (which is accessed in more generic dynamics) that is exponentially large in N . The model outlined in this appendix has only pairwise Majorana correlations, and a very restricted entanglement structure. However, this particular feature of the present problem is

likely to carry over to a larger set of models involving unitaries and measurements or projections for free fermions [53,156], since the key feature is having continuous, rather than discrete, replica symmetry [156,159].

APPENDIX C: CALCULATIONS FOR QUANTUM TREE

1. Numerical recursion for quantum tree

In this section we briefly describe the numerical procedure used to simulate the quantum tree. Due to the single-site Haar rotations the Schmidt basis of the tree Eq. (56) becomes uniformly distributed and we can thus characterize the wave function using only the two Schmidt values (that is, using a single real positive number between 0 and 1/2, which is the minimal Schmidt value squared $Z = \lambda_{\min}^2$).

The recursive procedure to generate the tree at generation $k + 1$ consists of using three singular values at generation k , Z_k , Z_k' , and Z_k'' and connecting them to a node using Eq. (57). Thus, the number of eigenvalues required to describe a certain instance of the tree *exactly* grows exponentially as 3^k , which is clearly not simulable at large k . Here we take advantage of the fact that in the case of the FMPT the nodes of the tree are statistically independent. Thus, at a certain level k we can generate a large constant pool of N singular values, where $1 \ll N \ll 3^k$. Assuming the pool spans the distribution function of Z_k faithfully we can then draw randomly three singular values from this pool to generate a member in the pool of the next generation. This is known as the ‘‘pool method’’ [105–107].

To verify that the pool spans the distribution of Z_k faithfully we test the convergence of the evolution of Z_k with the generation number k as a function of N . It is known that convergence in N can be very slow for the pool method [105,106]. For example, in Fig. 37 we present Z_k^{typ} at $\ln(r_c - r) = -5.3$ as a function of k for different values of N , which is the point closest to the critical point in Fig. 14. The origin of the strong N dependence lies in the exponent λ , which approaches 1/2 at the transition causing the distribution of Z to become broad. Upon tuning farther away from the critical point the minimal k and N required for convergence is found to decay rapidly (not shown in the figure).

Finally we also note that due to the forced measurements the distribution function of the singular values has a δ function at $Z = 0$ with a known prefactor. Namely, the probability of a singular value at generation k to be exactly zero is given by the recursive relation

$$f_k(p) = p + (1 - p)[f_{k-1}(p)]^3. \quad (\text{C1})$$

In principle we can keep these zeros in our pool. However, this is highly inefficient, especially when p starts

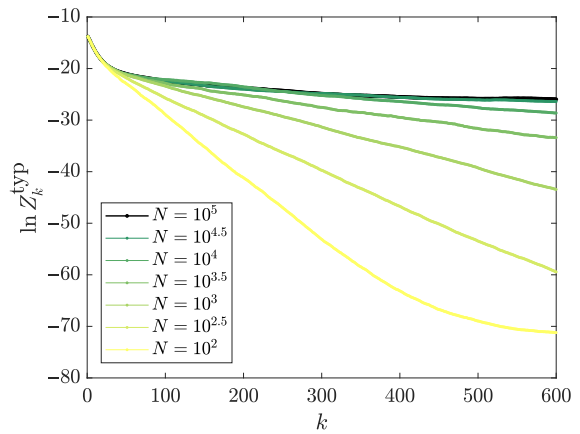


FIG. 37. Z_k^{typ} vs k at $\ln(r_c - r) = -5.3$ for Haar ensemble and for various pool sizes N . This plot demonstrates the convergence with pool size N and k for the point closest to the phase transition in Fig. 14. The minimal N and k required for convergence diminish for points farther away from the critical point. In the case of half-integer powers we round N to the closest integer.

to get close to the classical transition, where $f_\infty(p) = \left[\sqrt{1 + (2 - 3p)p/(1 - p)} - 1 \right] / 2$ becomes unity. Thus, in our simulations we keep only nonzero eigenvalues and, if needed, account for the zeros using Eq. (C1).

2. Averages of tree recursion constants

In Sec. E we describe the recursion relation for the singular values of the tree. The linearized recursion relation involves the random multiplicative constants in Eq. (E),

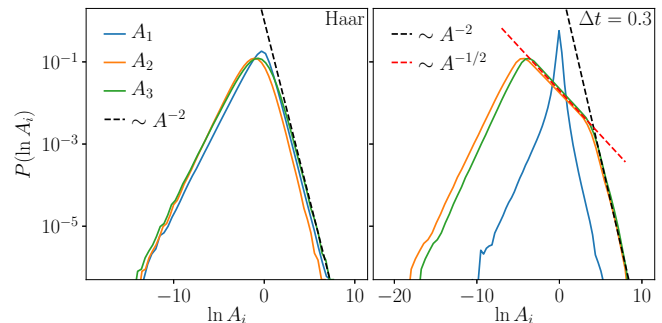


FIG. 38. The distribution function of the parameters $V_j \equiv \ln A_j$ for a Haar-random two-site unitary (left) and the ensemble in Eqs. (39) and (38) with $\Delta t = 0.3$. The data here is collected from $N = 10^7$ random unitaries.

which we repeat for convenience:

$$\begin{aligned} A_1 &= \frac{|U_{11}^{11}U_{21}^{21} - U_{21}^{11}U_{11}^{21}|^2}{(|U_{11}^{11}|^2 + |U_{21}^{21}|^2)^2}, \\ A_2 &= \frac{|U_{11}^{11}U_{12}^{21} - U_{12}^{11}U_{11}^{21}|^2}{(|U_{11}^{11}|^2 + |U_{21}^{21}|^2)^2}, \\ A_3 &= \frac{|U_{11}^{11}U_{11}^{22} - U_{11}^{12}U_{11}^{21}|^2}{(|U_{11}^{11}|^2 + |U_{21}^{21}|^2)^2}. \end{aligned} \quad (\text{C2})$$

In this appendix we derive the facts about the distribution of these quantities that are given in Sec. E. Some of these facts also hold for general node tensors, not necessarily expressed in terms of unitaries.

As in the text, we assume that the distribution of U is invariant under multiplication by single-site $U(2)$ matrices on any of its four legs. Initially however we do not assume that it is invariant under 2-site unitaries, i.e., we do not assume that U is Haar-distributed in $U(4)$.

First we show that, so long as the unitary U is non-trivially entangling (defined below) with probability 1, the average of any of the above quantities is exactly equal to 1,

$$\langle A_i \rangle = 1, \quad (\text{C3})$$

and also

$$\langle A_i^{1/2} \ln A_i \rangle = 0. \quad (\text{C4})$$

The argument is the same for any of the three A_i , so consider A_3 for definiteness. The argument only relies on the property of $U(2)$ invariance on a leg mentioned in Sec. B (together with the assumption that certain singular values are not fine tuned to zero), so they hold for more general choices of t satisfying this requirement. (We could also consider unitaries acting on more sites and trees with a larger branching number.)

The expression for A_3 involves only the matrix elements U_{11}^{ad} for $a, d = 1, 2$. Regarding this as a 2×2 matrix with row index a and column index d , we make a singular value decomposition, with positive singular values η_1 and η_2 :

$$U_{11}^{ad} = \sum_{\mu=1,2} w_{a\mu} \eta_{\mu} v_{\mu d}. \quad (\text{C5})$$

Here, w and v are $U(2)$ matrices. Now we note that, for any given U , the singular values of the tree we are considering are invariant under unitary basis transformations for the bond at the top of the tree. This implies that A_3 must be invariant if U_{11}^{ad} is multiplied by an arbitrary single-site unitary acting on the a index [see Eq. (41)]. We choose

this unitary to be the inverse of w , so that U_{11}^{ad} is replaced in Eq. (C2) by

$$U_{11}^{ad} \longrightarrow \eta_a v_{ad}. \quad (\text{C6})$$

Together with $|\det v|^2 = 1$, this gives the expression

$$A_3 = \frac{\eta_1^2 \eta_2^2}{(\eta_1^2 |v_{11}|^2 + \eta_2^2 |v_{21}|^2)^2}. \quad (\text{C7})$$

For some trivial, nonentangling two-site unitaries, such as the identity or swap, one of the singular values η is exactly zero, and A_3 vanishes. We assume that the distribution of U is such that, with probability 1, both singular values η are nonzero. This is our definition of “nontrivially entangling” above.

The above expression involves a single column, v_{a1} , of the $U(2)$ matrix v_{ad} . Since we assume that the distribution of U is invariant under single-site rotations, v_{ad} is Haar distributed, and v_{a1} is just a unit vector (with two complex or four real components) that must be averaged uniformly over the sphere S^3 .

This can be done in a standard way by relating the average over the sphere to a Gaussian average. Let us write the four real components of the unit vector (v_{11}, v_{21}) as $V = (w, x, y, z)$. If $\langle \dots \rangle_{\mu}$ is the Gaussian average with weight proportional to $e^{-\mu V^2}$, then

$$\langle \dots \rangle_{\mu} = 2\mu^2 \int dRR^3 e^{-\mu R^2} \langle \dots \rangle_{|V|=R}, \quad (\text{C8})$$

as we see by splitting the Gaussian integral on the lhs into radial and angular parts. The latter gives the integral over a sphere of fixed radius, which is the last expression on the rhs. We are interested in $\langle \dots \rangle_{|V|=1}$, the average over the unit sphere.

The Gaussian average of A_3 , which is of order $|V|^{-4}$, diverges at small R , so we instead first consider

$$f(\mu) \equiv \left\langle e^{-(\eta_1^2(w^2+x^2)+\eta_2^2(y^2+z^2))} \right\rangle_{\mu} = \frac{\mu^2}{(\mu + \eta_1^2)(\mu + \eta_2^2)}.$$

Using Eq. (C8) we may alternately write $f(\mu)$ as an average over the sphere. By scaling out a factor of R from the components of V we obtain an average over the unit sphere, and performing the R integral gives

$$f(\mu) = \left\langle \frac{\mu^2}{\{\mu + [\eta_1^2(w^2 + x^2) + \eta_2^2(y^2 + z^2)]\}^2} \right\rangle_{|V|=1}.$$

Equating the expressions for $f(\mu)$, and taking the limit $\mu \rightarrow 0$, the $U(2)$ Haar average is

$$\left\langle \frac{1}{(\eta_1^2 |v_{11}|^2 + \eta_2^2 |v_{21}|^2)^2} \right\rangle_v = \frac{1}{\eta_1^2 \eta_2^2}, \quad (\text{C9})$$

where we restore the previous notation for the vector. Plugging this into Eq. (C7) gives $\langle A_3 \rangle = 1$, as stated above, regardless of the precise distribution of η .

The same argument applies for A_1 and A_2 , using the appropriate singular value decomposition. Note that each of the A_i involves only a subset of the components of U , so that it effectively reduces to a 2×2 matrix, as above for the matrix U_{11}^{ad} . That is, in each case two of the four legs of U are set to index value “1.”

By multiplying $f(\mu)$ with μ^a and integrating over μ , we find

$$\langle A_i^\lambda \rangle = \frac{1}{2\lambda - 1} \left\langle \frac{1/H^{\lambda-1} - H^\lambda}{1 - H} \right\rangle, \quad (\text{C10})$$

with $H = \eta_1^2/\eta_2^2$. The remaining average on the rhs is over these singular values, again for the appropriate singular value decomposition of U .

Differentiating with respect to λ at $\lambda = 1/2$ gives Eq. (C4), irrespective of the distribution of H .

While Eqs. (C3) and (C4) simplified, more general moments depend on the detailed distribution of U . For the two-site Haar case we may write analytical formulas for $\langle A_i^\lambda \rangle$ [given in Eqs. (64) and (65) of the main text].

First consider A_3 , as above. This simplifies because, for a Haar-distributed unitary, U_{11}^{ad} can be viewed as a normalized and uniformly random vector in a Hilbert space of dimension 2×2 (a Page-random state). We are interested in the singular values when this state is split into two equal subsystems. Writing $s_i = \eta_i^2$, this distribution is [179],

$$\begin{aligned} P(s_1, s_2) &= 3\delta(s_1 + s_2 - 1)(s_1 - s_2)^2 \\ \rightarrow P(H) &= \frac{3(1 - H)^2}{(1 + H)^4}, \end{aligned}$$

for $0 < H < \infty$. Applying this to Eq. (C10) gives

$$\langle A_3^\lambda \rangle = \frac{\pi\lambda(1 - \lambda)}{\sin(\pi\lambda)}. \quad (\text{C11})$$

This is finite for $\lambda \in (-1, 2)$.

By the right-invariance of the Haar measure, the distribution of A_1 is the same as that of A_2 . We must consider the singular values for a decomposition of the matrix U_{b1}^{a1} , which is the upper left 2×2 block of a 4×4 Haar matrix. The distribution of singular values for such a sub-block of a Haar unitary may be found in Ref. [180]. Writing again $s_i = \eta_i^2$,

$$P(s_1, s_2) = 6(s_1 - s_2)^2, \quad (\text{C12})$$

with the constraint $0 < s_i < 1$ but, unlike in the previous case, no constraint on $s_1 + s_2$. Since there is a relabeling symmetry under $s_1 \leftrightarrow s_2$, or equivalently under $H \leftrightarrow$

$1/H$, we may insist $0 < H < 1$. Then

$$P(H) = 3(1 - H)^2. \quad (\text{C13})$$

Applying this to Eq. (C10) gives

$$\langle A_1^\lambda \rangle = \langle A_2^\lambda \rangle = \frac{12}{12 + 8\lambda - 7\lambda^2 - 2\lambda^3 + \lambda^4}, \quad (\text{C14})$$

as stated in the text. Again this is finite for $-1 < \lambda < 2$.

Finally, let us discuss the asymptotics of the distributions of the A_i . This will clarify the following point. In the main text we describe an ensemble of two-site unitaries with an entangling strength parameterized by Δt . In the limit of small Δt , these unitaries become closer and closer to the identity. For the identity, $A_2 = A_3 = 0$ exactly. But we show above that, for any nonzero value of Δt , no matter how small, $\langle A_2 \rangle = \langle A_3 \rangle = 1$. Therefore, the limit $\Delta t \rightarrow 0$ does not commute with the average over unitaries. This is because the distribution of A_i develops a long tail when Δt becomes small (see for example the intermediate tail in Fig. 38 for $\Delta t = 0.3$).

Define the two-component vectors

$$\phi_a = U_{11}^{a1} \quad \psi_a = U_{21}^{a1}, \quad \chi'_a = U_{11}^{a2} \quad \psi'_a = U_{12}^{a1}. \quad (\text{C15})$$

Then we may write Eqs. (C2) as

$$\begin{aligned} A_1 &= \frac{|\psi|^2|\phi|^2 - |\psi^\dagger\phi|^2}{|\phi|^4}, \\ A_2 &= \frac{|\psi'|^2|\phi|^2 - |\psi'^\dagger\phi|^2}{|\phi|^4}, \\ A_3 &= \frac{|\chi|^2|\phi|^2 - |\chi^\dagger\phi|^2}{|\phi|^4}. \end{aligned} \quad (\text{C16})$$

We see that A_i can become arbitrarily large if $|\phi|$ becomes small, with A_i scaling like $|\phi|^{-2}$ in this limit. Since ϕ has two complex (or four real) components, we expect that for a generic distribution of unitaries, the cumulative probability distribution of $|\phi|$ scales like $|\phi|^4$ at small $|\phi|$. This gives

$$P_{\ln}(\ln A_i) d \ln A_i \sim \frac{d \ln A_i}{A_i^2}, \quad A_i \gg 1 \quad (\text{C17})$$

at large A_i , as stated in the text. At small A_i , similar considerations for the numerators in Eq. (C2) show that generically

$$P(\ln A_i) d \ln A_i \sim A_i d \ln A_i, \quad A_i \ll 1. \quad (\text{C18})$$

These power laws are consistent with numerics and also with the fact that the moments $\langle A_i^\lambda \rangle$ for the Haar case diverge at $\lambda = 2$ and at $\lambda = -1$. The A_i are of course

correlated, but we do not consider their joint distribution here.

Now consider the case of a weakly entangling unitary with random single-site scramblers,

$$U = (u_1 \otimes u_2) e^{-i\Delta t H} (u_3 \otimes u_4), \quad (\text{C19})$$

with Δt small but fixed. (H may either be fixed, as in an ensemble discussed in the main text, or random.) We focus on A_2 and A_3 , whose distributions become broad at small Δt (that of A_1 does not). The two cases are similar, so consider A_3 , which is given by Eq. (C7). We expect that for small Δt we typically have $\eta_1^2 \sim \Delta t^2$ (here we keep only the scaling with Δt). Therefore, so long as $|v_{22}|^2 \gg \Delta t^2$, i.e., in the regime $A_3 \ll \Delta t^{-2}$, we have

$$A_3 = \frac{\eta_1^2 \eta_2^2}{(\eta_1^2 |v_{12}|^2 + \eta_2^2 |v_{22}|^2)^2} \sim \frac{\Delta t^2}{|v_{22}|^4}. \quad (\text{C20})$$

For small $|v_{22}|$, the cumulative distribution of $|v_{22}|$ scales like $|v_{22}|^2$, which gives

$$P(\ln A_3) d \ln A_3 \sim \frac{\Delta t}{\sqrt{A_3}} d \ln A_3, \quad \Delta t^2 \ll A_3 \ll \Delta t^{-2}. \quad (\text{C21})$$

On the other hand, for $A_3 \gg \Delta t^{-2}$, we expect to recover the generic exponent -2 [Eq. (C17)] for the distribution, suggesting

$$P(\ln A_3) d \ln A_3 \sim \frac{(\Delta t)^{-2}}{A_3^2} d \ln A_3, \quad A_3 \gg \Delta t^{-2}. \quad (\text{C22})$$

Similarly for $A_3 \ll \Delta t^2$ we expect to recover the exponent 1 for the generic case [Eq. (C18)]. The existence of three regimes with different power law exponents, 1, $-1/2$, and -2 , when Δt is relatively small, is in good agreement with the numerical data as shown in Fig. 38. Note that the scaling above, with the $-1/2$ tail being cut off at $A_3 \sim \Delta t^{-2}$, is also consistent with $\langle A_3 \rangle$ being of order 1 at small Δt .

3. Continuum recursion relation at $\Delta > 1$

We start with the FKPP equation with the spatially varying diffusion constant $D(x) = 1 + e^x$,

$$\partial_\tau H = \partial_x [D(x) \partial_x - a] H + \Delta H (1 - H), \quad (\text{C23})$$

in the regime $\Delta > 1$, such that the linearized problem is in the paramagnetic phase. The entanglement transition is at $a = -(\Delta + 1)$ so we write

$$a = -(\Delta + 1) + \sigma, \quad (\text{C24})$$

with $0 < \sigma \ll 1$. When $\sigma > 0$, we converge at late times to a stationary solution satisfying $\partial_\tau H = 0$. We would like

to determine the position x_f of the front in this solution, or equivalently the value of Z^{typ} (recall that $x_f \sim \ln Z^{\text{typ}}$).

Since H varies by an exponentially large factor over the relevant range of x , it is useful to look instead at the local exponential decay rate, which is order 1:

$$R(x) \equiv \partial_x \ln H(x). \quad (\text{C25})$$

From the stationary version of Eq. (C23), this satisfies

$$\begin{aligned} \partial_x R &= -\Delta f(x) - [1 - (\Delta - \sigma) f(x)] R - R^2 \\ &+ \Delta f(x) H(x), \end{aligned} \quad (\text{C26})$$

where

$$f(x) = \frac{1}{e^x + 1}. \quad (\text{C27})$$

R tends to 1 in the limit $x \rightarrow \infty$. [To see this note that in this limit Eq. (C23) becomes $\partial_x e^x \partial_x H = 0$. Together with $\lim_{x \rightarrow \infty} H = 0$ this gives $H \propto e^{-x}$.] Let us define

$$R(x) = -1 + \sigma S(x). \quad (\text{C28})$$

Then

$$\begin{aligned} \partial_x S &= -f(x) + [1 - (\Delta - \sigma) f(x)] S - \sigma S^2 \\ &+ \frac{\Delta}{\sigma} f(x) H(x). \end{aligned} \quad (\text{C29})$$

We will see in a moment that we must treat the cases $1 < \Delta < 2$ and $\Delta > 2$ separately. For now let us just note that if x is sufficiently large, the final term above will be subleading since both H and f tend to zero at large x . Assuming we are in the range of x where this term is negligible, and dropping terms that are subleading in σ ,

$$\partial_x S \simeq -f(x) + [1 - \Delta f(x)] S. \quad (\text{C30})$$

Solving this equation, and fixing the integration constant by demanding that S does not blow up as $x \rightarrow \infty$,

$$S = \frac{1}{\Delta(\Delta - 1)} \left[e^{-(\Delta-1)x} (e^x + 1)^\Delta - (e^x + \Delta) \right]. \quad (\text{C31})$$

Now consider this solution for large negative x

$$S = \frac{1}{\Delta(\Delta - 1)} \left[e^{(\Delta-1)|x|} + \dots \right]. \quad (\text{C32})$$

We check below that for $1 < \Delta < 2$, there is a range of x where this expansion is valid, i.e., where the final term in Eq. (C29) can indeed be neglected.

According to this expansion, as we increase $-x$, the value of R begins to increase significantly from -1 once

$$x \sim -\frac{1}{\Delta-1} \ln \frac{1}{\sigma}. \quad (\text{C33})$$

This suggests that the front is at $x_f \sim -\frac{1}{\Delta-1} \ln \frac{1}{\sigma}$, and that here we can match onto the stationary solution of the traveling-wave equation, Eq. (C23), with $\sigma = 0$ and with a spatially constant diffusion coefficient. (This matching makes sense since the forward part of this solution has $R = -1$ when $\sigma = 0$.)

However, we must check the self-consistency of our neglect of the final term in Eq. (C29). Assuming the above scaling for x_f , we have

$$\frac{1}{\sigma} f(x) H(x) \sim \sigma^{(2-\Delta)/(\Delta-1)} \frac{e^{-2x}}{1+e^{-x}} \quad (\text{C34})$$

for $x \gg x_f$. If $1 < \Delta < 2$, this term is indeed much smaller than the rhs of Eq. (C30) for $x \gg x_f$. Therefore, for this range of Δ the above analysis, giving

$$x_f \sim -\frac{1}{\Delta-1} \ln \frac{1}{\sigma} \quad (1 < \Delta < 2), \quad (\text{C35})$$

is self-consistent, though not rigorous.

On the other hand, for $\Delta > 2$, when the power of σ in Eq. (C34) is negative, this term cannot be dropped from Eq. (C29) for $x \gg x_f$. Numerically solving Eq. (C29) suggests that instead the final term in Eq. (C29) contributes at leading order if we fix x and take $\sigma \rightarrow 0$. Indeed, the alternative would be to have some x_c , with $x_f \ll x_c \ll 0$, such that the term is negligible for $x \gg x_c$ but not for $x_f < x < x_c$, and this may be seen to be inconsistent by examining the ratio of this term to the right-hand side of Eq. (C30). Using this fact, that H/σ should be of order 1 when $x \sim 0$, and assuming that $H \sim e^{-(x-x_f)}$ for $x \gg x_f$, we find that

$$x_f \sim -\ln \frac{1}{\sigma} \quad (\Delta > 2). \quad (\text{C36})$$

Equations (C35) and (C36) give the power laws $Z^{\text{typ}} \sim \sigma^{1/(\Delta-1)}$ and $Z^{\text{typ}} \sim \sigma$ stated in Sec. I.

These power laws can be checked directly by making a numerical solution of Eq. (C23) for $H(x, \tau)$ at different values of Δ and a . We use a numeric differential equation solver, solved over a wide domain of discrete values $x \in (x_L, x_R)$, with boundary conditions such that $H(x_L) = 1$ and $H(x_R) = 0$. These are exponentially close in x_L, x_R , respectively, to the true values for the solution on the infinite domain. An initial guess is used for $H(x, 0)$ and then evolved until a very long time $\tau = \tau_f$ in order to arrive at the steady-state solution, from which we can read off the position of the front. We define this as the value of

$x = x_{\text{front}}$ such that $H(x_{\text{front}}, \tau_f) = 1/2$. For the data presented in Fig. 17, $x_L = -70$, $x_R = 40$, $\tau_f = 10^8$, and the domain of x is discretized into 8001 points.

Making a linear fit of x_{front} against $\ln \sigma$ for a given value of Δ , and recalling $x_{\text{front}} \sim \ln Z^{\text{typ}}$, allows one to determine the value of the exponent γ , defined by $Z^{\text{yp}} \propto \sigma^\gamma$. We make this linear fit over the range of σ such that $-10 < \ln \sigma < -6$. Smaller σ requires very high numerical accuracy (a dense discretization of the domain of x), while at larger σ the critical behavior may not be apparent. The results are shown in the main text in Fig. 17. The error bars in this figure are defined by the difference in slope obtained from fits using only the left half of this range, $-10 < \ln \sigma < -8$, as compared to only the right half of the range, $-8 < \ln \sigma < -6$.

4. Minimal cut formula on tree

Assume that the minimal cut chops out m subtrees from the full tree, with each subtree being cut only once, at its apex. The singular values of a given subtree $a \in \{1, \dots, m\}$ are $\{\lambda_1^{(a)}, \lambda_2^{(a)}\}$, with $[\lambda_1^{(a)}]^2 + [\lambda_2^{(a)}]^2 = 1$ and $[\lambda_2^{(a)}]^2 = Z^{(a)}$. The full state may be written

$$|\psi\rangle = \sum_{i_1, \dots, i_m} \lambda_{i_1}^{(1)} \dots \lambda_{i_m}^{(m)} |i_1, \dots, i_m\rangle_{\text{subtrees}} |i_1, \dots, i_m\rangle_{\text{rest}}. \quad (\text{C37})$$

The states $|i_1, \dots, i_m\rangle_{\text{subtrees}}$ are products of the Schmidt states at the base of the subtrees, as in Sec. E, and are orthonormal. The states $|i_1, \dots, i_m\rangle_{\text{rest}}$, for the remaining spins, are neither normalized nor orthogonal. They are obtained by contracting the tensor network on the other side of the min cut with Schmidt states at the tops of the subtrees. The Rényi entropies are determined by the singular values of the (unnormalized) density matrix

$$\rho = \sum_{i_1, \dots, i_m} [\lambda_{i_1}^{(1)}]^2 \dots [\lambda_{i_m}^{(m)}]^2 |i_1, \dots, i_m\rangle_{\text{rest}} \langle i_1, \dots, i_m|_{\text{rest}}. \quad (\text{C38})$$

In the limit of small $Z^{(a)}$, with everything else fixed,

$$\rho \simeq |e_0\rangle \langle e_0| + \sum_{a=1}^m Z^{(a)} (|e_a\rangle \langle e_a| - |e_0\rangle \langle e_0|), \quad (\text{C39})$$

with $e_0 = |1, \dots, 1\rangle_{\text{rest}}$ and $e_a = |1, \dots, 2, \dots, 1\rangle_{\text{rest}}$, with the “2” in slot a . For $n > 1$ this gives

$$S_n \simeq \frac{n}{n-1} \sum_{a=1}^m \left(\frac{|e_a|^2 |e_0|^2 - |e_0^\dagger e_a|^2}{|e_0|^{2n}} \right) Z^{(a)} + \mathcal{O}(Z^2). \quad (\text{C40})$$

Each term is associated with a bond lying on the minimal cut. Let the height of this bond above the base be $k(a)$.

For a given term, the coefficient in brackets will vanish if the states e_a and e_0 become parallel. Each of these is a state in the “rest” Hilbert space, given by the tensor network made up of the tensors on one side of the minimal cut. The bonds lying on the minimal cut have fixed states attached to them: one of these is changed in going from e_0 to e_1 . Exploiting the fact that these truncated tensor networks are still trees, we can write e_0 and e_1 in terms of the singular value decompositions of two subtrees. We see that the term in brackets is of the same order as the singular value for a tree of depth $k(a)$. (This assumes that the singular values are not *growing* with k : this can occur if we fine tune the boundary conditions, but is not relevant to the case we are discussing.) Therefore, in the present limit of small Z we confirm the minimal cut conjecture in the main text, according to which each bond a on the minimal cut contributes an entanglement of order $Z^{(a)}Z'^{(a)}$, where $Z^{(a)}$ and $Z'^{(a)}$ are two random variables each distributed like Z_k for $k = k(a)$. (The $Z^{(a)}$ values are independent of each other and of the $Z'^{(a)}$ values, but the $Z'^{(a)}$ are correlated among themselves.)

APPENDIX D: MORE ON CIRCUIT SIMULATIONS

We present some additional details regarding the simulations of quantum circuits in Sec. V. The nonunitary time-evolution operator, V , for a given realization of the circuit is built by choosing at every step a measurement with probability r on a randomly chosen site or entangling two randomly chosen sites with the unitary with probability $1 - r$. For a system of size N , time progresses by one unit for every N unitaries applied. The Haar-random unitaries are generated using Mezzadri’s algorithm [181].

Note that the τ_{eff} described in the main text is obtained by taking a log derivative of $\mathcal{D}(t)$. Since, numerical derivatives are notoriously noisy, we smooth the data for $\mathcal{D}(t)$ using a Savitzky-Golay filter [182] and then take the derivative. Representative examples of τ_{eff} and the plateaux therein are shown in Figs. 39 and 40.

From the τ so obtained, we extract $a(r)$ by fitting the data to a form $a(r)N + b(r) + c(r)/N$. The fits are shown in Fig. 41.

APPENDIX E: TOY MODEL FOR THE SLOW DECAY OF S

In the entangled phase the operator entanglement has a plateau at an extensive value $S_n/N \simeq s_n > 0$, which persists for a time that scales exponentially in the number of spins, N . In this appendix we describe the crudest toy model for this plateau, which neglects both locality of the interactions and, for the most part, distinctions between different Rényi entropies. We consider the

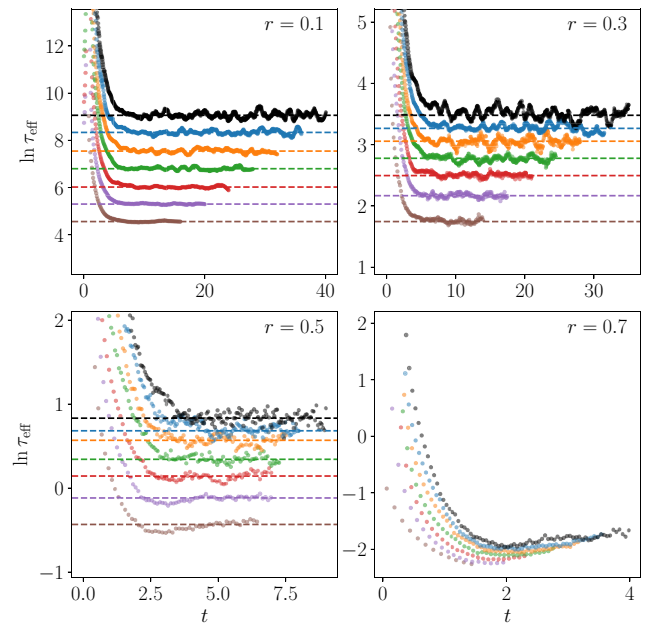


FIG. 39. τ_{eff} from time derivatives of $\ln \mathcal{D}^{\text{typ}}$. The plateaux denoted by the dashed lines show the $\tau(r, N)$ values, which are used to extract $a(r)$ (see Fig. 41). Different colors correspond to the different N following the same convention as in Fig. 22. Results are for the Haar circuit, with the top and bottom panels corresponding to the forced-measurement protocol and measurement protocol respectively.

forced-measurement case, where the circuit is made up of uncorrelated random pieces.

Imagine dividing up the circuit up into blocks of temporal duration Δt , which corresponds to writing the time-evolution operator $V(t)$ as a product of random matrices W_i , with each matrix of size $2^N \times 2^N$ and $i = 1, \dots, t/\Delta t$. Δt is chosen to be much larger than 1 but much smaller than the timescale τ that will emerge below. Each block has a singular value decomposition $W_i = U_i^{(1)} D U_i^{(2)\dagger}$. As a toy model, we treat the unitaries $U^{(1)}$ and $U^{(2)}$ as Haar random (neglecting locality) and we make the simplest choice of D that yields a given value of $S_1 = sN$, which is a flat

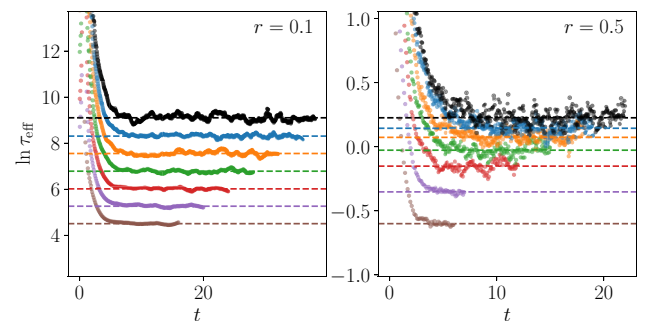


FIG. 40. Same as in Fig. 39 but for the Haar circuit with measurements.

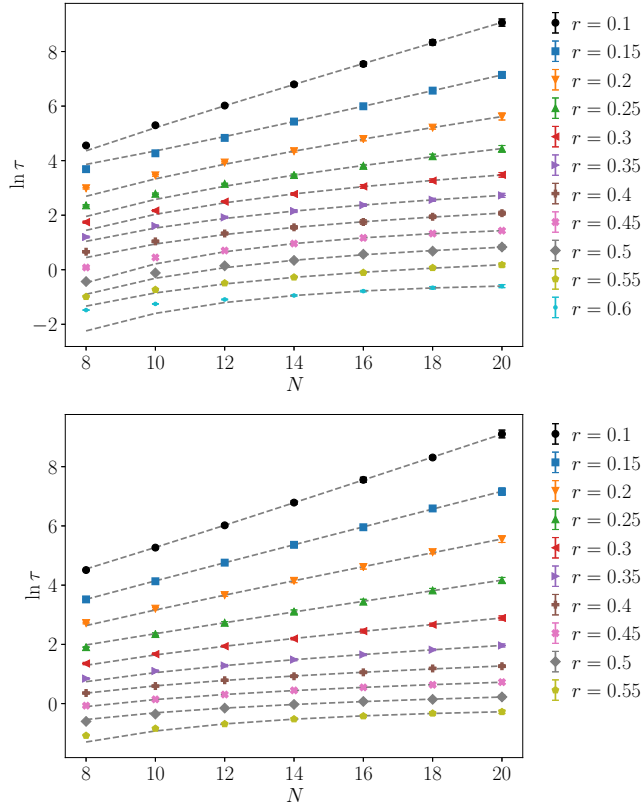


FIG. 41. Fits of $\ln \tau$ to a function of the form $a(r)N + b(r) + c(r)/N$ for different values of r for the Haar circuit with forced measurements (top) and measurements (bottom). We use $N \geq 14$ for the fits shown, which are used to extract $a(r)$ shown in Fig. 24. Fits including smaller N (not shown) are used to estimate the errorbars on $a(r)$.

entanglement spectrum:

$$D = \frac{1}{\sqrt{B}} \begin{pmatrix} \mathbb{I}_{B \times B} & 0 \\ 0 & 0 \end{pmatrix}, \quad B = e^{sN}, \quad 0 < s < \ln 2.$$

Note that the nonzero block is a small fraction of the size of the matrix for large N . Finally, we make an uncontrolled simplification by also treating the entanglement spectrum of $V(t)$ as flat:

$$V(t) = U_L(t)D(t)U_R(t)^\dagger, \\ D(t) = \frac{1}{\sqrt{B(t)}} \begin{pmatrix} \mathbb{I}_{B(t) \times B(t)} & 0 \\ 0 & 0 \end{pmatrix},$$

with $B(t) = e^{S(t)}$. Note that B and D refer to quantities for a single slice of width Δt , while $B(t)$ and $D(t)$ refer to the complete evolution operator up to time t .

We are interested in the singular values of the new evolution matrix $V(t + \Delta t)$, which may be written

$$V(t + \Delta t) = U'D(t)UDU''^\dagger \quad (\text{E1})$$

for Haar unitaries U' , U , U'' . These values are also the singular values of

$$\tilde{V} = D(t)UD. \quad (\text{E2})$$

Up to a normalization factor, \tilde{V} is just a rectangular block, of size $B(t) \times B$, taken from a Haar unitary of exponentially larger size ($2^N \times 2^N$). Correlations between unitary matrix elements become weaker as the size of the matrix increases, so we expect that we can treat them as Gaussian, with $\mathbb{E} U_{ab} = 0$ and $\mathbb{E} U_{ab}(U_{a'b'})^* = (1/2^N)\delta_{aa'}\delta_{bb'}$. (Higher cumulants are suppressed by powers of 2^N .) The singular values squared of \tilde{V} , denoted $v_i = \eta_i^2$, are eigenvalues of the $B(t) \times B(t)$ matrix

$$M = \tilde{V}\tilde{V}^\dagger. \quad (\text{E3})$$

When the matrix elements of \tilde{V} are complex Gaussian random numbers, this is as a Wishart random matrix (see, e.g., Ref. [179] for an application in a related context). The distribution of its eigenvalues depends on B as well as on $B(t)$. We assume that $1 \ll B(t) \ll B$. Normalizing the matrix so the s_i sum to one, the eigenvalue density for $\tilde{V}\tilde{V}^\dagger$ is the Marcenko-Pastur distribution,

$$\rho(v) = \frac{8\sqrt{v-v_-}\sqrt{v_+-v}}{\pi(v_+-v_-)^2v}, \quad v_{\pm} = \frac{[\sqrt{B} \pm \sqrt{B(t)}]^2}{BB(t)}.$$

From this distribution the operator Rényi entropies can be calculated as

$$e^{-(n-1)S_n(t+\Delta t)} = \int_{v_-}^{v_+} dv v^n \rho(v) \\ \simeq B(t)^{-(n-1)} \left[1 + \frac{n(n-1)}{2} \frac{B(t)}{B} + \dots \right] \\ S_n(t + \Delta t) - S(t) \simeq -\exp \left\{ - \left[sN - S(t) - \ln \frac{n}{2} \right] \right\}.$$

As expected, the entanglement spectrum does not remain flat. In our crude approximation, however, we neglect this, and apply the above transformation iteratively, so that in the continuum limit for times $\gg \Delta t$:

$$\partial_t S(t) \sim -C \exp\{-[sN - S(t)]\}. \quad (\text{E4})$$

Here C is an order-1 constant. At times larger than Δt , but short enough such that $S(t) \gg 1$ (which we assume above), this equation gives a solution:

$$S(t) = sN - \ln t + \dots \quad (\text{E5})$$

independently of the value of C . Note that we neglect random fluctuations (see Sec. I).

This analysis suggests a characteristic timescale τ with $\ln \tau \simeq sN$. We may verify this dependence directly in the opposite limit of asymptotically late times, where (as usual for a product of random matrices) there is a separation of scale between the largest singular value, the second largest, and so on. This separation allows us to consider only the two largest singular values. At a given time t , let them be normalized as

$$\{\eta_1, \eta_2\} = \{1, \epsilon(t)\}. \quad (\text{E6})$$

Then in place of M in Eq. (E3) we have a 2×2 matrix

$$M_{ik} = \epsilon^{i+k-2} \sum_{j=1}^B U_{ij} U_{kj}^*. \quad (\text{E7})$$

Each of the elements of M_{ik} is a different sum of many random variables, so we assume that M can be approximated as Gaussian:

$$\langle M_{ik} \rangle \simeq \frac{B\epsilon^{2(i-1)}\delta_{ik}}{2^N}, \quad \langle \langle M_{ik} M_{i'k'} \rangle \rangle \simeq \frac{B\epsilon^{2(i+i'-2)}\delta_{ik'}\delta_{i'k}}{2^{2N}}.$$

After absorbing a normalization constant into M ,

$$M = \begin{pmatrix} 1 & 0 \\ 0 & \epsilon^2 \end{pmatrix} + \frac{1}{\sqrt{B}} \begin{pmatrix} a & \epsilon\beta \\ \epsilon\beta^* & \epsilon^2 b \end{pmatrix}, \quad (\text{E8})$$

where a, b, β have mean zero and $\langle a^2 \rangle = \langle b^2 \rangle = \langle |\beta|^2 \rangle = 1$. The new (small) singular value squared is

$$\epsilon_{\text{new}}^2 = \epsilon^2 \left(1 - \frac{a-b}{\sqrt{B}} + \frac{a^2 - ab - |\beta|^2}{B} + \dots \right). \quad (\text{E9})$$

Let us study the typical value of this exponentially small quantity, defined by $(\epsilon_{\text{new}}^2)_{\text{typ}} = \exp(\ln \epsilon_{\text{new}}^2)$. We have

$$\ln \epsilon_{\text{new}}^2 = \ln \epsilon^2 + \frac{b-a}{\sqrt{B}} + \frac{a^2 - b^2 - 2|\beta|^2}{2B} + \dots \quad (\text{E10})$$

Here the average is taken over a, b, β . Note that the leading fluctuation term, of order $1/\sqrt{B}$, averages to zero. The next term, however, gives a negative drift under the recursion. Recalling that $B = e^{sN}$, and applying this map iteratively with each added block,

$$\epsilon_{\text{typ}}^2(t) \sim \exp\left(-\frac{t}{\tau}\right), \quad \tau = \Delta t \times \exp(sN). \quad (\text{E11})$$

Therefore, at the latest times,

$$S_n(t) \sim \exp\left(-\frac{t}{\tau}\right) \times \begin{cases} [n/(n-1)] & n > 1 \\ (2t/\tau) & n = 1 \end{cases}. \quad (\text{E12})$$

Thus, the main conclusion is that in this toy model sN sets both the value of the early time plateau and also the timescale for the late-time exponential decay, in agreement with the picture from the replica treatment.

APPENDIX F: FIELD THEORY: FURTHER DETAILS

1. $N \rightarrow 1$ ordered phase

In Sec. H we state that for $\mu^2 < 0$ the field theory of Sec. E has an ordered phase with $X_{ab} = f(\delta_{ab} - 1/N) + W_{ab}$, where $f = (-\mu^2/3g)[N/(N-2)]$ is the order parameter, and W_{ab} represents fluctuations around the saddle-point value whose quadratic Lagrangian is

$$\mathcal{L} = \frac{1}{2} \sum_{ab} (\partial W_{ab})^2 + \frac{-\mu^2 N}{2(2-N)} \left(\sum_{ab} W_{ab}^2 - 2 \sum_a W_{aa}^2 \right). \quad (\text{F1})$$

We want to check that if we compute the masses of the fluctuation modes using this expression, and then take $N \rightarrow 1$, these masses remain positive. Viewing W_{ab} as a vector, the term $\sum_a W_{aa}^2$ is $W.M.W$ where the matrix $M_{ab,cd}$ (with row index a, b and column index c, d) is 1 if $a = b = c = d$ and zero otherwise. We want the eigenvalues of M when projected onto the subspace of W satisfying $\sum_a W_{ab} = 0$ and $\sum_b W_{ab} = 0$, i.e., the eigenvalues of $\tilde{M}_{ab,cd} = P_{aa'} P_{bb'} M_{a'b',c'd'} P_{c'e} P_{d'e}$ where P is the projector $P_{aa'} = \delta_{aa'} - 1/N$. Since M is nonzero only when its four indices are equal, drawing a diagram shows that $\text{tr} \tilde{M}^k = \text{tr} S^k$, where $S_{a,b} = (1 - 2/N)\delta_{ab} + 1/N^2$. This gives the nonzero eigenvalues of \tilde{M} as $(N-1)/N \rightarrow 0$ with multiplicity 1 and $(N-2)/N \rightarrow -1$ with multiplicity $N-1$. Altogether, the eigenvalues of the matrix $(\mathbb{I} - 2\tilde{M})$, which appears in Eq. (F1) are either 1 or 3 in the limit and are positive.

2. Effect of YFY coupling when $N > 0$

Here we show that for $N > 0$ (for example, in the replica limit $N \rightarrow 1$, but *not* in the limit $N \rightarrow 0$), and in the vicinity of the critical point $r = 0$, the theory

$$\mathcal{L} = \sum_{ab} \left[\frac{1}{2} (\partial Y_{ab})^2 + r Y_{ab} + g Y_{ab}^3 \right] + \frac{m_F^2}{2} \sum_{abcd} Y_{ab} F_{ab,cd} Y_{cd} \quad (\text{F2})$$

can be reduced to the theory

$$\mathcal{L} = \sum_{ab} \left[\frac{1}{2} (\partial X_{ab})^2 + \frac{\mu^2}{2} X_{ab}^2 + g X_{ab}^3 \right], \quad (\text{F3})$$

for a matrix X with vanishing row and column sums, by discarding massive modes.

First, shifting the field by $Y_{ab} \rightarrow Y_{ab} + c$ with $c = -r/(2Nm_F^2) + \mathcal{O}(r^2)$ removes the linear term and generates a

mass term. The quadratic part of the Lagrangian is then

$$\mathcal{L}_2 = \frac{1}{2} \sum_{ab,cd} Y_{ab} [(k^2 + \mu^2)\delta_{ac}\delta_{bd} + m_F^2 F_{ab,cd}] Y_{cd}, \quad (\text{F4})$$

with $\mu^2 = -3gr/(Nm_F^2) + \mathcal{O}(r^2)$. As a matrix, $F = \mathbb{I} \otimes E + E \otimes \mathbb{I}$, where E is the $N \times N$ matrix with unit elements: $E_{ab} = 1$, so the matrix appearing in the brackets in Eq. (F4) is

$$(k^2 + \mu^2)\mathbb{I} \otimes \mathbb{I} + m_F^2 (\mathbb{I} \otimes E + E \otimes \mathbb{I}). \quad (\text{F5})$$

This can be decomposed in terms of the projection matrices $P_1 \equiv N^{-1}E$ and $P_{N-1} \equiv \mathbb{I} - P_1$ as

$$(k^2 + \mu^2) (P_{N-1} \otimes P_{N-1}) \quad (\text{F6})$$

$$+ (k^2 + \mu^2 + m_F^2 N) (P_1 \otimes P_{N-1} + P_{N-1} \otimes P_1) \quad (\text{F7})$$

$$+ (k^2 + \mu^2 + 2m_F^2 N) (P_1 \otimes P_1). \quad (\text{F8})$$

This is a decomposition into three representations of G_N of dimensions $(N-1)^2$, $2(N-1)$, and 1. We see that, at the critical point (where r and therefore μ^2 vanish) the second and the third representations remain massive and only the first becomes massless. Retaining only this representation is equivalent to fixing the row and column sums of Y to zero. Doing so and renaming the resulting field X yields precisely Eq. (F3). More precisely the two massive representations ought to be integrated out, renormalizing the values of the couplings in Eq. (F3).

These manipulations manifestly require $N > 0$: for example, they are appropriate for the replica limit $N \rightarrow 1$ relevant to the MPT. The replica limit $N \rightarrow 0$ must be handled separately.

3. Rewriting Lagrangian in $N \rightarrow 0$ limit

We give details of field redefinitions in Sec. G. First define the N -component vectors \mathbf{v}^+ , \mathbf{v}^- , and \mathbf{v}^i for $i = 2, \dots, N$ [133–136]:

$$\mathbf{v}^+ = \frac{1}{2} (1, 0, \dots, 0) + \frac{1}{2} \frac{(0, 1, \dots, 1)}{N-1}, \quad (\text{F9})$$

$$\mathbf{v}^- = \frac{1}{2} (1, 0, \dots, 0) - \frac{1}{2} \frac{(0, 1, \dots, 1)}{N-1}, \quad (\text{F10})$$

$$\mathbf{v}^i = (0, \dots, 0, 1, 0, \dots, 0) - \frac{(0, 1, \dots, 1)}{N-1}, \quad (\text{F11})$$

where the extra “1” in the third line is in the i th place. There are $N+1$ of these vectors but the \mathbf{v}^i are not linearly independent: $\sum_{i>1} \mathbf{v}^i = 0$.

We use these vectors to rewrite Y_{ab} in terms of $y_{\alpha\beta}$:

$$y_{\alpha\beta} = \mathbf{v}^\alpha \cdot Y \cdot \mathbf{v}^\beta = \sum_{ab} v_a^\alpha Y_{ab} v_b^\beta. \quad (\text{F12})$$

In this appendix we use a, b, a', b', \dots to denote indices that run from 1 to N , and α, β, \dots to denote indices that take the $N+1$ values $\{+, -, 2, \dots, N\}$. We use i, j, k to denote indices that run only over 2 to N . Note that

$$\sum_{j=2}^N y_{j,\beta} = 0, \quad \sum_{k=2}^N y_{\alpha,k} = 0. \quad (\text{F13})$$

(Alternately we could define an $N \times N$ matrix without this redundancy.)

To invert the transformation defining y , define $N+1$ -component vectors

$$\mathbf{x}^1 = (1, 1, 0, \dots, 0), \quad (\text{F14})$$

$$\mathbf{x}^{i>1} = (1, -1, 0, \dots, 1, \dots, 0). \quad (\text{F15})$$

The final 1 in the second line is in $(i+1)$ st place, which in our labeling convention corresponds to the component x_β^i with $\beta = i$. We have $\sum_\alpha x_\alpha^a v_a^\alpha = \delta_a^a$, so that

$$Y_{ab} = \mathbf{x}^a \cdot y \cdot \mathbf{x}^b = \sum_{\alpha\beta} x_\alpha^a y_{\alpha\beta} x_\beta^b. \quad (\text{F16})$$

Explicitly (for $j, k > 1$):

$$Y_{11} = (y_{++} + y_{+-} + y_{-+} + y_{--}), \quad (\text{F17})$$

$$Y_{1k} = (y_{++} - y_{+-} + y_{-+} - y_{--}) + (y_{+k} + y_{-k}), \quad (\text{F18})$$

$$Y_{j1} = (y_{++} + y_{+-} - y_{-+} - y_{--}) + (y_{j+} + y_{j-}), \quad (\text{F19})$$

$$Y_{jk} = (y_{++} - y_{+-} - y_{-+} + y_{--}) \quad (\text{F20})$$

$$+ (y_{j+} - y_{j-}) + (y_{+k} - y_{-k}) + y_{jk}. \quad (\text{F21})$$

Inserting these relations into the derivative term,

$$\begin{aligned} \sum_{ab} (\partial Y_{ab})^2 &= 8 (\partial y_{+-} \partial y_{-+} + \partial y_{++} \partial y_{--}) \\ &+ 4 \sum_j \partial y_{j+} \partial y_{j-} \\ &+ 4 \sum_k \partial y_{+k} \partial y_{-k} \\ &+ \sum_{jk} (\partial y_{jk})^2 \\ &+ \mathcal{O}(N), \end{aligned} \quad (\text{F22})$$

where the final line contains terms whose coefficients contain an explicit factor of N , which we assume can be neglected in the limit $N \rightarrow 0$.

Next consider the F term in the Lagrangian, which has the form

$$\sum_{abcd} Y_{ab} F_{ab,cd} Y_{cd} = \sum_a \left(\sum_b Y_{ab} \right)^2 + \sum_b \left(\sum_a Y_{ab} \right)^2. \quad (\text{F23})$$

This simplifies to

$$\begin{aligned} \sum_{abcd} Y_{ab} F_{ab,cd} Y_{cd} = & 16y_{--} (y_{+-} + y_{-+}) \\ & + 4 \sum_j y_{j-}^2 \\ & + 4 \sum_k y_{-k}^2 \\ & + \mathcal{O}(N). \end{aligned} \quad (\text{F24})$$

The linear term in the Lagrangian is simply

$$\sum_{ab} Y_{ab} = 4y_{--} + \mathcal{O}(n). \quad (\text{F25})$$

Since the coefficient of the linear term is zero at the critical point, we assign “engineering” dimensions to the fields such that the quadratic terms in Eqs. (F22) and (F24) are marginal. Denoting the space-time dimension by D , the inverse length dimensions of the fields are, in order of increasing scaling dimension,

$$[y_{++}] = (D - 6)/2, \quad (\text{F26})$$

$$[y_{j+}] = [y_{+k}] = (D - 4)/2, \quad (\text{F27})$$

$$[y_{+-}] = [y_{-+}] = [y_{jk}] = (D - 2)/2, \quad (\text{F28})$$

$$[y_{j-}] = [y_{-k}] = (D + 0)/2, \quad (\text{F29})$$

$$[y_{--}] = (D + 2)/2, \quad (\text{F30})$$

so that if the number of “+” indices for a component of y is n_+ , and the number of “-” indices is n_- , the engineering dimension is

$$x(n_+, n_-) = \frac{(D - 2) - 2(n_+ - n_-)}{2}. \quad (\text{F31})$$

This formula implies that, at a given order in y , terms with the largest number of + indices and the smallest number of - indices are most relevant.

Now consider the additional terms in the Lagrangian beyond those in Eqs. (F22), (F25), (F24), order by order in Y .

G_N symmetry does not allow any linear terms other than Eq. (F25). At quadratic order the remaining possibilities are $\sum_{ab} Y_{ab}^2$ and $(\sum_{ab} Y_{ab})^2$. The former is redundant—it can be cancelled by a shift in Y because of the presence of $\sum_{ab} Y_{ab}^3$. We also check this below in the new parameterization. The latter, $(\sum_{ab} Y_{ab})^2$, is proportional to $(y_{--})^2$ by Eq. (F25), so is less relevant than the terms on the rhs of Eq. (F24) and can be neglected.

Cubic terms are obtained by contracting indices in

$$Y_{ab} Y_{a'b'} Y_{a''b''}, \quad (\text{F32})$$

i.e., by setting some indices equal to others and summing. Left indices may only be set equal to other left indices, and similarly for right indices. This allows three different types of index contraction: an index (e.g., a) can be summed without being set equal to any other index; two indices can be set equal and then summed (e.g., $a = a'$); or three indices can be set equal and summed ($a = a' = a''$). When we rewrite the contracted expression in terms of $y_{\alpha\beta} y_{\alpha'\beta'} y_{\alpha''\beta''}$, the single, double, and triple index contractions lead to contractions with the tensors

$$d_{\alpha}^{(1)} = \sum_a x_{\alpha}^a, \quad d_{\alpha\alpha'}^{(2)} = \sum_a x_{\alpha}^a x_{\alpha'}^a, \quad d_{\alpha\alpha'\alpha''}^{(3)} = \sum_a x_{\alpha}^a x_{\alpha'}^a x_{\alpha''}^a, \quad (\text{F33})$$

respectively, for the indices $\alpha, \alpha', \alpha'' \in \{+, -, 2, \dots, N\}$. For each of these, we may check the maximal value of

$$\Delta \equiv n_+ - n_-, \quad (\text{F34})$$

the difference in the number of + and - indices, that may appear on the right-hand side. (Δ should not be confused with a scaling dimension.) Since the difference in the total number of + and - indices is what determines the engineering dimension of a field or a product of fields [Eq. (F31)], identifying Δ_{\max} for each type of index contraction allows us to say which types of contraction will give the most relevant cubic terms: they are those for which the sum of Δ_{\max} , over all contractions, is largest.

Explicitly,

$$d_{\alpha}^{(1)} = (N, 2 - N, 1, \dots, 1)_{\alpha}. \quad (\text{F35})$$

However, y vanishes when contracted with $(0, 1, \dots, 1)$ [Eq. (F13)]. After dropping this part, and taking the limit

$N \rightarrow 0$ directly in the coefficients,

$$d_{\alpha}^{(1)} \rightarrow 2\delta_{\alpha,-}. \quad (\text{F36})$$

Therefore, this pattern of index contraction contributes $\Delta = -1$. Using similar simplifications,

$$d_{\alpha,\alpha'}^{(2)} \rightarrow 2(\delta_{+\alpha}\delta_{-\alpha'} + \delta_{-\alpha}\delta_{+\alpha'}) + \sum_{j=2}^N \delta_{j\alpha}\delta_{j\alpha'}. \quad (\text{F37})$$

This pattern of index contraction contributes $\Delta = 0$. Finally, $d^{(3)}$ contains various patterns of index contraction with different values for Δ :

$$d_{\alpha,\alpha',\alpha''}^{(3)} \rightarrow 2(\delta_{-\alpha}\delta_{+\alpha'}\delta_{+\alpha''} + \dots) \quad (\text{F38})$$

$$+ \sum_{i=2}^N (\delta_{+\alpha}\delta_{i\alpha'}\delta_{i\alpha''} + \dots) \quad (\text{F39})$$

$$+ \sum_i \delta_{i\alpha}\delta_{i\alpha'}\delta_{i\alpha''} \quad (\text{F40})$$

$$- \sum_i (\delta_{-\alpha}\delta_{i\alpha'}\delta_{i\alpha''} + \dots) \quad (\text{F41})$$

$$+ 2\delta_{-\alpha}\delta_{-\alpha'}\delta_{-\alpha''}. \quad (\text{F42})$$

(Ellipses indicate terms related to those shown by cyclic permutations of $\alpha, \alpha', \alpha''$.) The first two lines on the rhs have $\Delta = 1$; all the others have smaller values of Δ . If we are interested only in keeping the most relevant terms in a given expression, we can truncate to only the first two lines:

$$\tilde{d}_{\alpha,\alpha',\alpha''}^{(3)} \equiv 2(\delta_{-\alpha}\delta_{+\alpha'}\delta_{+\alpha''} + \dots) \quad (\text{F43})$$

$$+ \sum_{i=2}^N (\delta_{+\alpha}\delta_{i\alpha'}\delta_{i\alpha''} + \dots). \quad (\text{F44})$$

Therefore, the most relevant cubic terms allowed by G_N symmetry are those with $\Delta = 2$ (in total, i.e., counting both row and column indices) that arise by discarding the less-relevant parts of $\sum_{ab} Y_{ab}^3$:

$$\sum_{ab} Y_{ab}^3 \rightarrow \sum \tilde{d}_{\alpha,\alpha',\alpha''}^{(3)} \tilde{d}_{\beta,\beta',\beta''}^{(3)} Y_{\alpha\beta} Y_{\alpha'\beta'} Y_{\alpha''\beta''}, \quad (\text{F45})$$

which is equal to

$$\begin{aligned} \left(\sum_{ab} Y_{ab}^3 \right)_{\Delta=2} &= 12y_{++} (y_{++}y_{--} + 2y_{+-}y_{-+}) \\ &+ 12y_{++} \left(y_{-k}y_{+k} + y_{j-}y_{j+} + \frac{y_{jk}y_{jk}}{4} \right) \\ &+ 6(y_{+-}y_{j+}y_{j+} + y_{-+}y_{+k}y_{+k}) \\ &+ 6y_{j+}y_{+k}y_{jk}. \end{aligned} \quad (\text{F46})$$

(Repeated j or k indices are summed from 2 to N .) All other contractions of YYY , such as $\sum_{aa'bb'} Y_{ab}Y_{ab'}Y_{a'b'}$, give terms that are strictly less relevant according to the engineering dimensions.

Let us check that, having dropped less relevant terms from the Lagrangian in the process of rewriting it in terms of y , the coupling of the quadratic term

$$\begin{aligned} \frac{1}{4} \sum_{ab} Y_{ab}^2 &= 2(y_{+-}y_{-+} + y_{++}y_{--}) + y_{j+}y_{j-} + y_{+k}y_{-k} \\ &+ \frac{y_{jk}y_{jk}}{4}, \end{aligned} \quad (\text{F47})$$

is still redundant. The shift $Y_{ab} \rightarrow Y_{ab} + c$ for each element corresponds to shifting $y_{++} \rightarrow y_{++} + c$ and leaving other elements of y unchanged [by Eq. (F12)]. Under this shift, the cubic term, Eq. (F46), generates precisely Eq. (F47), with a coefficient of order c [and the linear term in Eq. (F25) with a coefficient of order c^2]. Therefore, by an appropriate shift of y_{++} we may eliminate the quadratic term, Eq. (F47).

-
- [1] T. Basché, S. Kummer, and C. Bräuchle, Direct spectroscopic observation of quantum jumps of a single molecule, *Nature* **373**, 132 (1995).
 - [2] S. Gleyzes, S. Kuhr, C. Guerlin, J. Bernu, S. Deléglise, U. Busk Hoff, M. Brune, J.-M. Raimond, and S. Haroche, Quantum jumps of light recording the birth and death of a photon in a cavity, *Nature* **446**, 297 (2007).
 - [3] R. Vijay, D. H. Slichter, and I. Siddiqi, Observation of Quantum Jumps in a Superconducting Artificial Atom, *Phys. Rev. Lett.* **106**, 110502 (2011).
 - [4] L. Robledo, L. Childress, H. Bernien, B. Hensen, P. F. A. Alkemade, and R. Hanson, High-fidelity projective readout of a solid-state spin quantum register, *Nature* **477**, 574 (2011).
 - [5] Z. K. Mineev, S. O. Mundhada, S. Shankar, P. Reinhold, R. Gutiérrez-Jáuregui, R. J. Schoelkopf, M. Mirrahimi, H. J. Carmichael, and M. H. Devoret, To catch and reverse a quantum jump mid-flight, *Nature* **570**, 200 (2019).
 - [6] B. Skinner, J. Ruhman, and A. Nahum, Measurement-Induced Phase Transitions in the Dynamics of Entanglement, *Phys. Rev. X* **9**, 031009 (2019).
 - [7] Y. Li, X. Chen, and M. P. A. Fisher, Quantum Zeno effect and the many-body entanglement transition, *Phys. Rev. B* **98**, 205136 (2018).
 - [8] A. Chan, R. M. Nandkishore, M. Pretko, and G. Smith, Unitary-projective entanglement dynamics, *Phys. Rev. B* **99**, 224307 (2019).
 - [9] Y. Li, X. Chen, and M. P. A. Fisher, Measurement-driven entanglement transition in hybrid quantum circuits, *Phys. Rev. B* **100**, 134306 (2019).
 - [10] M. Szyniszewski, A. Romito, and H. Schomerus, Entanglement transition from variable-strength weak measurements, *Phys. Rev. B* **100**, 064204 (2019).

- [11] S. Choi, Y. Bao, X.-L. Qi, and E. Altman, Quantum Error Correction in Scrambling Dynamics and Measurement-Induced Phase Transition, *Phys. Rev. Lett.* **125**, 030505 (2020).
- [12] M. J. Gullans and D. A. Huse, Dynamical purification phase transitions induced by quantum measurements, [arXiv:1905.05195](https://arxiv.org/abs/1905.05195) [quant-ph] (2019).
- [13] Y. Bao, S. Choi, and E. Altman, Theory of the phase transition in random unitary circuits with measurements, *Phys. Rev. B* **101**, 104301 (2020).
- [14] C.-M. Jian, Y.-Z. You, R. Vasseur, and A. W. W. Ludwig, Measurement-induced criticality in random quantum circuits, *Phys. Rev. B* **101**, 104302 (2020).
- [15] Y. Li, X. Chen, A. W. W. Ludwig, and M. P. A. Fisher, Conformal invariance and quantum non-locality in hybrid quantum circuits, [arXiv:2003.12721](https://arxiv.org/abs/2003.12721) [quant-ph] (2020).
- [16] A. Zabalo, M. J. Gullans, J. H. Wilson, S. Gopalakrishnan, D. A. Huse, and J. H. Pixley, Critical properties of the measurement-induced transition in random quantum circuits, *Phys. Rev. B* **101**, 060301 (2020).
- [17] M. J. Gullans and D. A. Huse, Scalable Probes of Measurement-Induced Criticality, *Phys. Rev. Lett.* **125**, 070606 (2020).
- [18] Q. Tang and W. Zhu, Measurement-induced phase transition: A case study in the nonintegrable model by density-matrix renormalization group calculations, *Phys. Rev. Res.* **2**, 013022 (2020).
- [19] Y. Fuji and Y. Ashida, Measurement-induced quantum criticality under continuous monitoring, *Phys. Rev. B* **102**, 054302 (2020).
- [20] O. Lunt and A. Pal, Measurement-induced entanglement transitions in many-body localized systems, [arXiv:2005.13603](https://arxiv.org/abs/2005.13603) [quant-ph] (2020).
- [21] M. Sznyszewski, A. Romito, and H. Schomerus, Universality of entanglement transitions from stroboscopic to continuous measurements, [arXiv:2005.01863](https://arxiv.org/abs/2005.01863) [cond-mat.stat-mech] (2020).
- [22] X. Turkeshi, R. Fazio, and M. Dalmonte, Measurement-induced criticality in $(2 + 1)$ -dimensional hybrid quantum circuits, *Phys. Rev. B* **102**, 014315 (2020).
- [23] M. V. Regemortel, Z.-P. Cian, A. Seif, H. Dehghani, and M. Hafezi, Entanglement entropy scaling transition under competing monitoring protocols, [arXiv:2008.08619](https://arxiv.org/abs/2008.08619) [quant-ph] (2020).
- [24] R. Fan, S. Vijay, A. Vishwanath, and Y.-Z. You, Self-organized error correction in random unitary circuits with measurement, [arXiv:2002.12385](https://arxiv.org/abs/2002.12385) [cond-mat.stat-mech] (2020).
- [25] O. Shtanko, Y. A. Kharkov, L. P. García-Pintos, and A. V. Gorshkov, Classical models of entanglement in monitored random circuits, [arXiv:2004.06736](https://arxiv.org/abs/2004.06736) [cond-mat.dis-nn] (2020).
- [26] S. Vijay, Measurement-driven phase transition within a volume-law entangled phase, [arXiv:2005.03052](https://arxiv.org/abs/2005.03052) [quant-ph] (2020).
- [27] Y. Li and M. P. A. Fisher, Statistical Mechanics of Quantum Error-Correcting Codes, [arXiv:2007.03822](https://arxiv.org/abs/2007.03822) [quant-ph] (2020).
- [28] D. Aharonov, Quantum to classical phase transition in noisy quantum computers, *Phys. Rev. A* **62**, 62311 (2000).
- [29] D. Gottesman, An Introduction to Quantum Error Correction and Fault-Tolerant Quantum Computation, [arXiv:0904.2557](https://arxiv.org/abs/0904.2557) [quant-ph] (2009).
- [30] L. Bonnes and A. M. Läuchli, Superoperators vs. Trajectories for matrix product state simulations of open quantum system: A case study, [arXiv:1411.4831](https://arxiv.org/abs/1411.4831) [cond-mat.quant-gas] (2014).
- [31] J. Napp, R. L. L. Placa, A. M. Dalzell, F. G. S. L. Brandao, and A. W. Harrow, Efficient classical simulation of random shallow 2D quantum circuits, [arXiv:2001.00021](https://arxiv.org/abs/2001.00021) [quant-ph] (2019).
- [32] M. B. Plenio and P. L. Knight, The quantum-jump approach to dissipative dynamics in quantum optics, *Rev. Mod. Phys.* **70**, 101 (1998).
- [33] A. J. Daley, Quantum trajectories and open many-body quantum systems, *Adv. Phys.* **63**, 77 (2014).
- [34] T. Zhou and A. Nahum, Emergent statistical mechanics of entanglement in random unitary circuits, *Phys. Rev. B* **99**, 174205 (2019).
- [35] P. Hayden, S. Nezami, X.-L. Qi, N. Thomas, M. Walter, and Z. Yang, Holographic duality from random tensor networks, *J. High Energy Phys.* **2016**, 1 (2016).
- [36] R. Vasseur, A. C. Potter, Y.-Z. You, and A. W. W. Ludwig, Entanglement transitions from holographic random tensor networks, *Phys. Rev. B* **100**, 134203 (2019).
- [37] A. Nahum and B. Skinner, Entanglement and dynamics of diffusion-annihilation processes with majorana defects, *Phys. Rev. Res.* **2**, 023288 (2020).
- [38] S. Sang and T. H. Hsieh, Measurement protected quantum phases, [arXiv:2004.09509](https://arxiv.org/abs/2004.09509) [cond-mat.stat-mech] (2020).
- [39] A. Lavasani, Y. Alavirad, and M. Barkeshli, Measurement-induced topological entanglement transitions in symmetric random quantum circuits, [arXiv:2004.07243](https://arxiv.org/abs/2004.07243) [quant-ph] (2020).
- [40] M. Ippoliti, M. J. Gullans, S. Gopalakrishnan, D. A. Huse, and V. Khemani, Entanglement phase transitions in measurement-only dynamics, [arXiv:2004.09560](https://arxiv.org/abs/2004.09560) [quant-ph] (2020).
- [41] N. Lang and H. P. Büchler, Entanglement Transition in the Projective Transverse Field Ising Model, [arXiv:2006.09748](https://arxiv.org/abs/2006.09748) [cond-mat.dis-nn] (2020).
- [42] A. Nahum, S. Vijay, and J. Haah, Operator Spreading in Random Unitary Circuits, *Phys. Rev. X* **8**, 021014 (2018).
- [43] T. Zhou and A. Nahum, Entanglement Membrane in Chaotic Many-Body Systems, *Phys. Rev. X* **10**, 031066 (2020).
- [44] N. Hunter-Jones, Unitary designs from statistical mechanics in random quantum circuits, [arXiv:1905.12053](https://arxiv.org/abs/1905.12053) [quant-ph] (2019).
- [45] H. Liu and S. Vardhan, Entanglement entropies of equilibrated pure states in quantum many-body systems and gravity, [arXiv:2008.01089](https://arxiv.org/abs/2008.01089) [hep-th] (2020).
- [46] P. Zanardi, Entanglement of quantum evolutions, *Phys. Rev. A* **63**, 40304 (2001).
- [47] T. Prosen and I. Pižorn, Operator space entanglement entropy in a transverse Ising chain, *Phys. Rev. A* **76**, 032316 (2007).
- [48] I. Pižorn and T. Prosen, Operator space entanglement entropy in xy spin chains, *Phys. Rev. B* **79**, 184416 (2009).

- [49] 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* **50**, 234001 (2017).
- [50] T. Zhou and D. J. Luitz, Operator entanglement entropy of the time evolution operator in chaotic systems, *Phys. Rev. B* **95**, 094206 (2017).
- [51] C. Jonay, D. A. Huse, and A. Nahum, Coarse-grained dynamics of operator and state entanglement, [arXiv:1803.00089](https://arxiv.org/abs/1803.00089) [cond-mat.stat-mech] (2018).
- [52] X. Cao, A. Tilloy, and A. D. Luca, Entanglement in a fermion chain under continuous monitoring, *SciPost Phys.* **7**, 24 (2019).
- [53] X. Chen, Y. Li, M. P. A. Fisher, and A. Lucas, Emergent conformal symmetry in nonunitary random dynamics of free fermions, *Phys. Rev. Res.* **2**, 033017 (2020).
- [54] O. Alberton, M. Buchhold, and S. Diehl, Trajectory dependent entanglement transition in a free fermion chain – from extended criticality to area law, [arXiv:2005.09722](https://arxiv.org/abs/2005.09722) [cond-mat.stat-mech] (2020).
- [55] In general the membrane tension can depend on the Rényi index n [34]. It can also depend on the orientation of the membrane [51], but here we are interested only in membranes that are “horizontal” on large scales.
- [56] L. Fidkowski, J. Haah, and M. B. Hastings, How dynamical quantum memories forget, [arXiv:2008.10611](https://arxiv.org/abs/2008.10611) [quant-ph] (2020).
- [57] M. J. Gullans and D. A. Huse (Private communication).
- [58] B. Derrida and H. Spohn, Polymers on disordered trees, spin glasses, and traveling waves, *J. Stat. Phys.* **51**, 817 (1988).
- [59] Y.-Y. Shi, L.-M. Duan, and G. Vidal, Classical simulation of quantum many-body systems with a tree tensor network, *Phys. Rev. A* **74**, 022320 (2006).
- [60] L. Tagliacozzo, G. Evenbly, and G. Vidal, Simulation of two-dimensional quantum systems using a tree tensor network that exploits the entropic area law, *Phys. Rev. B* **80**, 235127 (2009).
- [61] V. Murg, F. Verstraete, O. Legeza, and R. M. Noack, Simulating strongly correlated quantum systems with tree tensor networks, *Phys. Rev. B* **82**, 205105 (2010).
- [62] P. Silvi, V. Giovannetti, S. Montangero, M. Rizzi, J. I. Cirac, and R. Fazio, Homogeneous binary trees as ground states of quantum critical hamiltonians, *Phys. Rev. A* **81**, 062335 (2010).
- [63] W. Li, J. von Delft, and T. Xiang, Efficient simulation of infinite tree tensor network states on the bethe lattice, *Phys. Rev. B* **86**, 195137 (2012).
- [64] N. Nakatani and G. K.-L. Chan, Efficient tree tensor network states (TTNS) for quantum chemistry: Generalizations of the density matrix renormalization group algorithm, *J. Chem. Phys.* **138**, 134113 (2013).
- [65] V. Murg, F. Verstraete, R. Schneider, P. R. Nagy, and O. Legeza, Tree tensor network state with variable tensor order: An efficient multireference method for strongly correlated systems, *J. Chem. Theory Comput.* **11**, 1027 (2015).
- [66] G. Vidal, Entanglement Renormalization, *Phys. Rev. Lett.* **99**, 220405 (2007).
- [67] B. Swingle, Entanglement renormalization and holography, *Phys. Rev. D* **86**, 065007 (2012).
- [68] J. Lopez-Piqueres, B. Ware, and R. Vasseur, Mean-field entanglement transitions in random tree tensor networks, *Phys. Rev. B* **102**, 064202 (2020).
- [69] R. N. C. Pfeifer, G. Evenbly, and G. Vidal, Entanglement renormalization, scale invariance, and quantum criticality, *Phys. Rev. A* **79**, 040301 (2009).
- [70] The term “random tensor network” allows for almost any structure, so infinite numbers of universality classes can in principle be accessed, most of them extremely fine tuned.
- [71] K. H. Fischer and J. A. Hertz, *Spin Glasses* (Cambridge University Press, Cambridge, UK, 1993), Vol. 1.
- [72] For generic versions of the MPT and FMPT, the emergent space-time symmetry is of course partly a conjecture. It is perhaps made more plausible by the existence of such symmetries in some simpler limiting models. The minimal cut problem [6], and also some alternative $q \rightarrow \infty$ limits [13,14], map to percolation problems, which have this symmetry. Some measurement induced critical points with a free-fermion structure also map to conformally invariant models [37–39,41,54]. Conformal invariance in $1 + 1D$ Clifford measurement circuits has been demonstrated numerically [9,15]. There is numerical evidence that the dynamical exponent is unity for the Haar-random MPT [6]. Finally, we may use dual-unitary circuits to set up measurement circuits that have 90° rotational invariance in space time even microscopically (we discuss this elsewhere).
- [73] S_0 counts the (logarithm of the) number of nonzero Schmidt values in the singular value decomposition of a state or, as we focus on here, an operator.
- [74] F. Pastawski, B. Yoshida, D. Harlow, and J. Preskill, Holographic quantum error-correcting codes: Toy models for the bulk/boundary correspondence, *J. High Energy Phys.* **2015**, 149 (2015).
- [75] H. Casini, H. Liu, and M. Mezei, Spread of entanglement and causality, *J. High Energy Phys.* **2016**, 77 (2016).
- [76] A. Nahum, J. Ruhman, S. Vijay, and J. Haah, Quantum Entanglement Growth under Random Unitary Dynamics, *Phys. Rev. X* **7**, 031016 (2017).
- [77] J. T. Chayes, L. Chayes, and R. Durrett, Critical behavior of the two-dimensional first passage time, *J. Stat. Phys.* **45**, 933 (1986).
- [78] B. Bollobás, *Random Graphs*, Cambridge Studies in Advanced Mathematics (Cambridge University Press, 2001), 2nd ed.
- [79] C. M. Fortuin and P. W. Kasteleyn, On the random-cluster model: I. Introduction and relation to other models, *Physica* **57**, 536 (1972).
- [80] R. K. P. Zia and D. J. Wallace, Critical behaviour of the continuous n -component Potts model, *J. Phys. A* **8**, 1495 (1975).
- [81] D. J. Amit, Renormalization of the Potts model, *J. Phys. A* **9**, 1441 (1976).
- [82] J. Cardy, *Scaling and Renormalization in Statistical Physics*, Cambridge Lecture Notes in Physics (Cambridge University Press, 1996).

- [83] A. Aharony, Y. Gefen, and A. Kapitulnik, Scaling at the percolation threshold above six dimensions, *J. Phys. A* **17**, L197 (1984).
- [84] J. L. Cardy, Critical percolation in finite geometries, *J. Phys. A* **25**, L201 (1992).
- [85] A conclusive numerical check of the exponent in Eq. (27) would require larger system sizes since it requires the scaling function in Fig. 6 to have large negative argument.
- [86] We write this formula for the case where $r_c - r$ is of order 1. Otherwise the lower limit on the range may involve a critical timescale that is larger than 1 but much smaller than τ .
- [87] For a naive picture of the scaling of S_0 with N , consider a minimal cut that has zero temporal width within the entangled phase. There are $\mathcal{O}(t)$ choices for the time at which to place the cut. Each choice has a random cost, which for the present illustration we assume to be Gaussian with variance N , arising from a sum of $\mathcal{O}(N)$ random contributions. Taking the minimum gives the formula above. The second term is subleading so long as $\ln t \ll N$.
- [88] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, *Introduction to Algorithms* (MIT Press, Cambridge, MA, USA, 2009).
- [89] As usual, this geometrical disconnection is reflected in the vanishing of the zeroth Rényi entropy S_0 between apex and base.
- [90] This is a weaker condition than that satisfied by the Haar ensemble of two-site unitaries, which is invariant under left or right multiplication by any two-site unitary.
- [91] This is because when we redraw the section of the circuit as a tree, we turn some of the unitaries upside down (which amounts to taking a transpose) or reflect them left-right. The discrete invariances of the distributions mean that we do not need to keep track of this.
- [92] The link may contain multiple projections, but this reduces to the case with a single projection. In the case where all the projection operators are identical, this is immediate. More generally, it holds because the distribution of unitaries is invariant under single-site rotations (and because the overall normalization factor for the tree is not important).
- [93] In linear algebra terms, the tree represents a $2 \times 2^\ell$ matrix: 2 is the number of states associated with the bond at the apex, and 2^ℓ is the number of states associated with the ℓ spins at the base. We are interested in the singular values in the SVD for this matrix.
- [94] The trees we encounter in the circuit have a nonzero probability of terminating before they achieve k generations, in which case Z_k is identically zero. So long as we are below the classical transition, an order-1 fraction of the probability distribution for Z_k is supported on nonzero values even when k is finite but large. For other tree tensor networks Z_k may be nonzero with probability 1 (for any finite k), in which case we simply define $\ln Z_k^{\text{typ}} = \langle \ln Z_k \rangle$.
- [95] Considering $T \equiv T_{k+1}$ as a state, then $\lambda_1^2 + \lambda_2^2 = 1$ and $\lambda_1^4 + \lambda_2^4 = \text{Tr } \rho^2 / (\text{Tr } \rho)^2$, where ρ is the un-normalized reduced density matrix for the spin at the apex: $\rho_{a,a'} = \sum_{bcd} T_{bcd}^a T_{bcd}^{*a'}$.
- [96] This can be seen from the fact that Z_{k+1} is manifestly independent of Z_k' and Z_k'' in this limit.
- [97] The latter converges at large k to a constant below 1, so long as we are in the classically percolating phase for the tree. The value of this plateau can be set to zero by a linear transformation of G . See Sec. H.
- [98] R. A. Fisher, The wave of advance of advantageous genes, *Ann. Eugen.* **7**, 355 (1937).
- [99] The value $\lambda = 1$ appears here because it is the exponential decay constant for the initial condition at $k = 0$: $G_0(x) \sim 1 - \langle Z_0 \rangle e^{-x}$ at large x . At late times the solution converges to a traveling wave in which the decay constant is $\min\{1, \lambda_*\}$. That is, the decay constant is either that of the initial condition, or that of the minimal-speed wave, whichever decays more slowly as $x \rightarrow \infty$.
- [100] Recall that we assume various invariances of the distribution of unitaries to simplify the treatment (Sec. B).
- [101] Equation (83) is the exact growth rate of $\langle Z_k \rangle$ for any p in the linear problem, but it is only for $p < p_{\text{glass}}$ that Z_k^{typ} has the same growth rate as $\langle Z_k \rangle$. It is c_p as defined by Z_k^{typ} that will be relevant when we include nonlinearity. The failure of the annealed approximation when $\lambda_* < 1$ is because, in this regime the distribution of Z becomes broad in the sense that $\lim_{k \rightarrow \infty} \langle Z/Z^{\text{typ}} \rangle = \infty$. [In this regime the tail in the probability distribution for $\ln(Z/Z^{\text{typ}})$ decays as $e^{-\lambda_* \ln(Z/Z^{\text{typ}})}$.]
- [102] These have $Z = 0$, so do not contribute to the recursion.
- [103] This is because, close to the classical transition, only a small fraction, of order $(2/3 - p)$, of subtrees survive. As a result most nodes in a “pruned” tree have a single descendant, with a fraction of order $(2/3 - p)^{m-1}$ having m descendants ($m \leq 3$).
- [104] We note that due to the forced measurements there is a finite probability for Z to be exactly zero, i.e., the distribution function has a δ function with a finite weight. These are instances where the tree is classically disconnected. When we compute Z^{typ} and present distribution functions we do not include these trivial instances.
- [105] J. D. Miller and B. Derrida, Weak-disorder expansion for the Anderson model on a tree, *J. Stat. Phys.* **75**, 357 (1994).
- [106] C. Monthus and T. Garel, Anderson transition on the Cayley tree as a traveling wave critical point for various probability distributions, *J. Phys. A* **42**, 75002 (2008).
- [107] I. García-Mata, O. Giraud, B. Georgeot, J. Martin, R. Dubertrand, and G. Lemarié, Scaling Theory of the Anderson Transition in Random Graphs: Ergodicity and Universality, *Phys. Rev. Lett.* **118**, 166801 (2017).
- [108] The multiplicative factor appearing here is the probability that a tree never terminates. In the limit of large k this is also the probability that Z is nonzero.
- [109] E. Brunet and B. Derrida, Shift in the velocity of a front due to a cutoff, *Phys. Rev. E* **56**, 2597 (1997).
- [110] By examining the first-order equation satisfied by $R = \partial_x \ln H$ after dropping the $\mathcal{O}(H^2)$ and $\mathcal{O}(\sigma)$ terms, we see that the value of this slope at some arbitrarily chosen value x in the vicinity of the origin must be tuned to the correct $\mathcal{O}(1)$ value, R_x , in order to match onto the correct solution at large positive x , where the slope is -1 . Further this value R_x is necessarily less than $-\sqrt{\Delta}$, which is the approximate slope of Eq. (96) in the region where the sine is of order 1. To achieve this, we must take advantage

- of the negative divergence of the cotangent in Eq. (97) at argument π . This means that to leading order in σ we must have $\phi = \pi$.
- [111] Defining $\mathbf{t}_{b_1 \dots b_\ell}$ as the vector with components $t_{b_1 \dots b_\ell}^a$, then $A_1 =_{\text{law}} |\mathbf{t}_{11 \dots 1}|^{-4} \left(|\mathbf{t}_{21 \dots 1}|^2 |\mathbf{t}_{11 \dots 1}|^2 - |\mathbf{t}_{21 \dots 1}^\dagger \mathbf{t}_{11 \dots 1}|^2 \right) = |\mathbf{t}_{11 \dots 1}|^{-4} |t_{21 \dots 1}^1 t_{11 \dots 1}^2 - t_{11 \dots 1}^1 t_{21 \dots 1}^2|^2$ (for A_m the “2” is the m th subscript). The argument in Appendix 2 making use of the invariance of t under single-leg rotations also applies here.
- [112] The prefactor depends on the singular vectors at the top of the subtrees, but is of order 1.
- [113] Here we use the U(2) invariance property of the unitaries to show that lower terms vanished by phase cancellation.
- [114] We use a naive linear extrapolation: the detailed functional form of the subleading corrections to S/N may depend on the precise regime of N and t (Sec. I).
- [115] Note that the latter value of r is right at the putative critical point. This value is chosen to avoid being above, or to close to, the classical critical point $r = 0.8$ where the network is trivially disconnected.
- [116] These finite size effects, together with the fact that extracting $a(r)$ requires two separate fits [to go from \mathcal{D}_{typ} to $\tau(r, N)$ and then to $a(r)$] also make it hard to estimate error bars on $a(r)$ accurately.
- [117] S. J. Garratt and J. T. Chalker, Many-body quantum chaos and the local pairing of feynman histories, [arXiv:2008.01697](https://arxiv.org/abs/2008.01697) [cond-mat.stat-mech] (2020).
- [118] J. Cardy, Logarithmic conformal field theories as limits of ordinary CFTs and some physical applications, *J. Phys. A* **46**, 494001 (2013).
- [119] A. Nahum, Field theories for measurement and entanglement phase transitions (to be published).
- [120] Simplifications arise in the fully unitary case. Even there, for general N it is possible to have negative Boltzmann weights W_σ . Simplifications also arise at large local Hilbert space dimension [13,14,34,36].
- [121] The dependence is only on the cycle structure of $\sigma^{-1}\tau$ thanks to G_N symmetry.
- [122] M. Stone and P. Goldbart, *Mathematics for Physics: A Guided Tour for Graduate Students* (Cambridge University Press, Cambridge, UK, 2009).
- [123] A given realization has no symmetry. Because of the distribution we choose for the disorder there are statistical symmetries, i.e., symmetries of the disorder distribution. For example the distribution is invariant under $S \rightarrow -S$ on a given site. However, this symmetry does not act non-trivially on the effective field X , so we expect it could be broken without changing the universality class of the transition under discussion.
- [124] We use the set of N vectors $\mathbf{e}^1, \dots, \mathbf{e}^N$, each of $N-1$ components, that are familiar in the context of the Potts field theory [80,81] (see Appendix A 2) and satisfy $\sum_a e_a^\alpha e_a^\beta = \delta_{\alpha\beta} \cdot X_{ab} = \sum_{\alpha, \beta=1}^{N-1} (e_a^\alpha e_a^\beta) \phi_{\alpha\beta}$.
- [125] E. Pytte and J. Rudnick, Scaling, equation of state, and the instability of the spin-glass phase, *Phys. Rev. B* **19**, 3603 (1979).
- [126] J. R. L. de Almeida and D. J. Thouless, Stability of the sherrington-kirkpatrick solution of a spin glass model, *J. Phys. A: Math. General* **11**, 983 (1978).
- [127] G. Parisi, A sequence of approximated solutions to the S-K model for spin glasses, *J. Phys. A: Math. General* **13**, L115 (1980).
- [128] J. Cardy, Logarithmic correlations in quenched random magnets and polymers, [arXiv:cond-mat/9911024](https://arxiv.org/abs/cond-mat/9911024) [cond-mat.stat-mech] (1999).
- [129] R. Vasseur, J. L. Jacobsen, and H. Saleur, Logarithmic observables in critical percolation, *J. Stat. Mech.* **2012**, L07001 (2012).
- [130] R. Vasseur and J. L. Jacobsen, Operator content of the critical Potts model in d dimensions and logarithmic correlations, *Nucl. Phys. B* **880**, 435 (2014).
- [131] At first sight the interpretation of these additional fields may seem obscure, given that for a permutation matrix they are trivial constants. This may be more transparent in the approach of Sec. VID.
- [132] M. E. Fisher, Yang-Lee Edge Singularity and ϕ^3 Field Theory, *Phys. Rev. Lett.* **40**, 1610 (1978).
- [133] J. L. Cardy, Nonperturbative aspects of supersymmetry in statistical mechanics, *Phys. D* **15**, 123 (1985).
- [134] J. L. Cardy and A. J. McKane, Field theoretic approach to the study of Yang-Lee and Griffiths singularities in the randomly diluted Ising model, *Nucl. Phys. B* **257**, 383 (1985).
- [135] J. Cardy, Exact scaling functions for self-avoiding loops and branched polymers, *J. Phys. A* **34**, L665 (2001).
- [136] A. Kaviraj, S. Rychkov, and E. Trevisani, Random field Ising model and Parisi-Sourlas supersymmetry. Part I. Supersymmetric CFT, *J. High Energy Phys.* **2020**, 1 (2020).
- [137] G. Parisi and N. Sourlas, Random Magnetic Fields, Supersymmetry, and Negative Dimensions, *Phys. Rev. Lett.* **43**, 744 (1979).
- [138] A. J. Bray and M. A. Moore, Scaling theory of the random-field Ising model, *J. Phys. C: Solid State Phys.* **18**, L927 (1985).
- [139] T. C. Lubensky and J. Isaacson, Field Theory for the Statistics of Branched Polymers, Gelation, and Vulcanization, *Phys. Rev. Lett.* **41**, 829 (1978).
- [140] G. Parisi and N. Sourlas, Critical Behavior of Branched Polymers and the Lee-Yang Edge Singularity, *Phys. Rev. Lett.* **46**, 871 (1981).
- [141] D. C. Brydges and J. Z. Imbrie, Branched polymers and dimensional reduction, *Ann. Math.* **1019** (2003).
- [142] At the quadratic level, these various sectors can be replaced by a theory with fermions but without a replica limit [133,137] (for example, $N-2 \rightarrow -2$ copies of a bosonic theory can be replaced by a fermionic version of the theory).
- [143] A. Chan, A. De Luca, and J. T. Chalker, Spectral Statistics in Spatially Extended Chaotic Quantum Many-Body Systems, *Phys. Rev. Lett.* **121**, 060601 (2018).
- [144] This is similar to what happens for the classical minimal cut. If η may be treated as Gaussian, the correction to the entanglement is of order $\sqrt{N} \ln t$ in this regime.
- [145] N. Hunter-Jones, Operator growth in random quantum circuits with symmetry, [arXiv:1812.08219](https://arxiv.org/abs/1812.08219) [quant-ph] (2018).

- [146] V. L. Nguen, B. Z. Spivak, and B. I. Shklovskii, Tunnel hopping in disordered systems, *Sov. Phys. JETP* **89**, 1770 (1985).
- [147] Y.-C. Zhang, Directed polymers with complex amplitudes, *Europhys. Lett.* **9**, 113 (1989).
- [148] E. Medina, M. Kardar, Y. Shapir, and X. R. Wang, Interference of Directed Paths in Disordered Systems, *Phys. Rev. Lett.* **62**, 941 (1989).
- [149] J. Cook and B. Derrida, Lyapunov exponents of large, sparse random matrices and the problem of directed polymers with complex random weights, *J. Stat. Phys.* **61**, 961 (1990).
- [150] Y. Y. Goldschmidt and T. Blum, Directed walks with complex random weights: Phase diagram and replica symmetry breaking, *J. Phys.* **12**, 1607 (1992).
- [151] B. Derrida, M. R. Evans, and E. R. Speer, Mean field theory of directed polymers with random complex weights, *Comm. Math. Phys.* **156**, 221 (1993).
- [152] D. Bernard and T. Jin, Open Quantum Symmetric Simple Exclusion Process, *Phys. Rev. Lett.* **123**, 080601 (2019).
- [153] D. Bernard and T. Jin, Solution to the Quantum Symmetric Simple Exclusion Process: The Continuous Case, [arXiv:2006.12222](https://arxiv.org/abs/2006.12222) [math-ph] (2020).
- [154] R. Frassek, C. Giardina, and J. Kurchan, Duality in quantum transport models, [arXiv:2008.03476](https://arxiv.org/abs/2008.03476) [cond-mat.stat-mech] (2020).
- [155] F. Thiel, E. Barkai, and D. A. Kessler, First Detected Arrival of a Quantum Walker on an Infinite Line, *Phys. Rev. Lett.* **120**, 040502 (2018).
- [156] T. Swann, J. T. Chalker, and A. Nahum, (to be published).
- [157] A. Chan and J. T. Chalker, Loop soups in quasi-one dimension: The Anderson tower of states and the Poisson-Dirichlet distribution (to be published).
- [158] C. W. J. Beenakker, Random-matrix theory of quantum transport, *Rev. Mod. Phys.* **69**, 731 (1997).
- [159] F. Evers and A. D. Mirlin, Anderson transitions, *Rev. Mod. Phys.* **80**, 1355 (2008).
- [160] Formally, when $b' > b$, we use as our modes the coefficients of the right eigenvectors of T , which is equivalent to analytically continuing the formula for the $b' < b$ case. The square root nonanalyticity does not appear in the physical quantities below, only in the intermediate formulas.
- [161] A. Theumann, Bond percolation problem in a semi-infinite medium. Landau-Ginzburg theory, *Phys. Rev. B* **19**, 6295 (1979).
- [162] J. P. Carton, Surface effects and percolation: The repulsive case, *J. Phys. Lett.* **41**, 175 (1980).
- [163] K. De'Bell and J. W. Essam, Mean field theory of percolation with application to surface effects, *J. Phys. A* **14**, 1993 (1981).
- [164] Translating these BCs into the continuum field theory Eq. (A30) gives boundary magnetic field terms in Eq. (A23) of the form $h\delta(t)e_\sigma\cdot\phi(t)$ (and similarly at the final time boundary) with $h \propto N^{1/2}$. At first sight one worries that the N dependence of h introduces another exponent that could appear in scaling forms. However, we believe that the basic point is just that h diverges with N , so that the asymptotic scaling forms are those of the $h \rightarrow \infty$ limit and the detailed N dependence of h determines only subleading corrections.
- [165] A. Coniglio, in *Physics of Finely Divided Matter: Proceedings of the Winter School, Les Houches, France, March 25–April 5, 1985*, edited by N. Boccara and M. Daoud (Springer Science & Business Media, 2012), Vol. 5.
- [166] M. Aizenman, On the number of incipient spanning clusters, *Nucl. Phys. B* **485**, 551 (1997).
- [167] L. de Arcangelis, Multiplicity of infinite clusters in percolation above six dimensions, *J. Phys. A* **20**, 3057 (1987).
- [168] S. Fortunato, D. Stauffer, and A. Coniglio, Percolation in high dimensions is not understood, *Phys. A* **334**, 307 (2004).
- [169] S. Fortunato, A. Aharony, A. Coniglio, and D. Stauffer, Number of spanning clusters at the high-dimensional percolation thresholds, *Phys. Rev. E* **70**, 056116 (2004).
- [170] R. Kenna and B. Berche, Universal finite-size scaling for percolation theory in high dimensions, *J. Phys. A* **50**, 235001 (2017).
- [171] In the saddle-point language this factor would come from a fluctuation integral (cf. e.g., [183,184]): it is consistent with $(Q-2) \rightarrow -1$ modes with eigenvalue $\propto |\delta r|$.
- [172] S_0 for a single spin is either 0 or 1 bits, with the former holding if the spin is connected to no other spins on the final time boundary. This may be written as a one-point function of the spin operator in a system where other boundary spins are fixed. I_0 for two spins is either 0, 1, or 2 bits, and is given by a minimal cut formula. If we assume that the probabilities for both $I_0 = 1$ and $I_0 = 2$ have the same scaling form (this can be demonstrated in $1+1D$ [6]) then we can focus on the simpler case $I_0 = 2$, which occurs only if the two spins are connected to each other, but not to any other spins on the final time boundary. This is the two-point function of the same Potts operator.
- [173] T. C. Lubensky and M. H. Rubin, Critical phenomena in semi-infinite systems. II. Mean-field theory, *Phys. Rev. B* **12**, 3885 (1975).
- [174] A. J. Bray and M. A. Moore, Critical behaviour of semi-infinite systems, *J. Phys. A* **10**, 1927 (1977).
- [175] C. Candu, J. L. Jacobsen, N. Read, and H. Saleur, Universality classes of polymer melts and conformal sigma models, *J. Phys. A* **43**, 142001 (2010).
- [176] A. Nahum, J. T. Chalker, P. Serna, M. Ortuño, and A. M. Somoza, Phase transitions in three-dimensional loop models and the cP^{n-1} sigma model, *Phys. Rev. B* **88**, 134411 (2013).
- [177] A. Nahum and J. T. Chalker, Universal statistics of vortex lines, *Phys. Rev. E* **85**, 031141 (2012).
- [178] A. Nahum, P. Serna, A. M. Somoza, and M. Ortuño, Loop models with crossings, *Phys. Rev. B* **87**, 184204 (2013).
- [179] C. Nadal, S. N. Majumdar, and M. Vergassola, Statistical distribution of quantum entanglement for a random bipartite state, *J. Stat. Phys.* **142**, 403 (2011).
- [180] M. Kieburg, A. B. J. Kuijlaars, and D. Stivigny, Singular value statistics of matrix products with truncated unitary matrices, *Int. Math. Res. Not.* **2016**, 3392 (2015).

- [181] F. Mezzadri, How to generate random matrices from the classical compact groups, [arXiv:math-ph/0609050](https://arxiv.org/abs/math-ph/0609050) [math-ph] (2006).
- [182] A. Savitzky and M. J. E. Golay, Smoothing and differentiation of data by simplified least squares procedures, *Anal. Chem.* **36**, 1627 (1964).
- [183] J. L. Cardy, Electron localization in disordered systems and classical solutions in Ginzburg-Landau field theory, *J. Phys. C* **11**, L321 (1978).
- [184] R. Hayn and W. John, Instanton approach to the conductivity of a disordered solid, *Nucl. Phys. B* **348**, 766 (1991).