We formulate a theory of the many-body localization transition based on a novel real space renormalization group (RG) approach. The results of this theory are corroborated and intuitively explained with a phenomenological effective description of the critical point and of the “badly conducting” state found near the critical point on the delocalized side. The theory leads to the following sharp predictions: (i) The delocalized state established near the transition is a Griffiths phase, which exhibits sub-diffusive transport of conserved quantities and sub-ballistic spreading of entanglement. The anomalous diffusion exponent $\alpha < 1/2$ vanishes continuously at the critical point. The system does thermalize in this Griffiths phase. (ii) The many-body localization transition is controlled by a new kind of infinite randomness RG fixed point, where the broadly distributed scaling variable is closely related to the eigenstate entanglement entropy. Dynamically, the entanglement grows as $\sim \log t$ at the critical point, as it also does in the localized phase. (iii) In the vicinity of the critical point the ratio of the entanglement entropy to the thermal entropy, and its variance (and in fact all moments) are scaling functions of $L/\xi$, where $L$ is the length of the system and $\xi$ is the correlation length, which has a power-law divergence at the critical point.

I. INTRODUCTION

Anderson had postulated, already in his original paper on localization, that closed many-body systems undergoing time evolution would not come to thermal equilibrium if subject to sufficiently strong randomness [1]. Significant theoretical effort has been devoted in the last few years to understand this phenomenon, the only known generic exception to thermalization (see e.g. [2, 8] for recent reviews). The recent work led to classification of many-body localization (MBL) as a distinct dynamical phase of matter, characterized by a remarkable set of defining properties: (i) there are locally accessible observables that do not relax to their equilibrium values and hence can be related to a set of quasi-local integrals of motion [4–8]; (ii) even after arbitrarily long time evolution retrievable quantum information persists in the system and may be extracted from local degrees of freedom [9, 10]; (iii) entanglement entropy grows with time evolution only as a logarithmic function of time [11–13].

In spite of the progress in understanding the MBL phase, very little is known about the dynamical phase transition which separates it from the delocalized thermal phase. Part of the difficulty lies in the fundamental difference between the energy eigenstates found on either side of the transition. Eigenstates in the thermal phase are expected to obey the eigenstate thermalization hypothesis, which, in particular, implies extensive (i.e. volume law) entanglement entropy. The non-locality of quantum mechanics is fully exploited in such states, where information resides in highly non local entities: the exponentially many expansion coefficients of the wave-function in terms of local basis states. On the other hand, in the many-body localized phase the eigenstates feature area-law entanglement entropy akin to quantum ground states. Hence a putative critical point separating the two types of states would be unlike any other known phase transition. Ground state quantum critical points and dynamical critical points which occur inside the localized phase mark transitions between area-law states, whereas thermal critical points are transitions between distinct states with extensive (i.e. volume law) entropy. The need to describe this critical point, where the eigenstates change from area law to volume law entanglement, and hence the quantum information in some sense escapes from localized degrees of freedom to highly non-local ones, calls for a new theoretical approach.

In this paper we develop a strong disorder renormalization group framework which can address this many-body localization phase transition. We find a transition controlled by an infinite randomness RG fixed point, where the broad distributions are of a scaling variable directly related to the entanglement entropy of the system’s eigenstates. Thus, using this RG scheme we obtain finite size scaling results for the probability distribution of the entanglement entropy near this phase transition. A corollary of the analysis is that the phases adjacent to this critical point are Griffiths phases, where some properties are dominated by rare regions. On the delocalized side of the transition, there is a thermal Griffiths phase showing anomalous (sub-diffusive) transport and sub-linear entanglement growth under time evolution, due to rare, locally insulating regions that impede the transport but do not prevent thermalization.

Before proceeding, we mention the relation to recent work on the MBL transition. Most of this work has relied on exact diagonalization of very small systems [14–16]. In particular, the numerical results of Ref. [15] suggested an infinite randomness critical point. More recently, Kjall et. al. [16] identified a peak in the variance of the eigenstate entanglement entropy as a sensitive variable for locating and characterizing the transition. This indeed turns out to be related to the main scaling variable in our theory. Finally, recent work identified and
explored the sub-diffusive regime in the vicinity of the transition [17, 18]. In this paper we present a comprehensive theory which naturally explains and unifies the different phenomena associated with dynamics and entanglement near the many-body localization phase transition.

II. RG SCHEME

A strong disorder renormalization group scheme has been developed recently to describe the dynamics within the many-body localized phase [3, 19, 20]. This approach, however, neglects resonances, i.e. non-local modes involving many of the microscopic degrees of freedom. Because these are the very processes that lead to delocalization, a new approach, which incorporates the physics of resonances, is needed in order to describe the many-body localization transition.

The microscopic systems we have in mind include disordered spin chains as well as interacting lattice particles hopping in a random potential in one dimension. But in order to capture the effect of resonances we will concede the fully microscopic starting point and instead work with an effective coarse grained model of the system, which we expect nonetheless provides a faithful description of the system near the critical point. We are able to consider within the same framework closed systems with energy conservation as well as periodically driven (Floquet) systems, which lack any conservation laws other than the unitarity and locality of their time evolution. We work at the energy density that corresponds to infinity near the critical point. Note that the Hilbert space dimension $N$ of the coarse grained model of a chain of $L$ microscopic spins-1/2 is still $2^L$, exactly as the bare model. Thus we do not “integrate out” states. However, the retained information is reduced because we now keep only a few parameters for each block $i$ of $l_i$ spins. In the course of renormalization, pairs of adjacent blocks are joined into longer blocks, so the total number of retained parameters is steadily reduced.

To identify the parameter $g_i$ for a given block $i$ it is useful to consider the time and energy scales that characterize the block. Of course each block is characterized by a typical many-body level spacing $\Delta_i = W 2^{-i} \sqrt{l_i}$, where $W$ is a microscopic bare energy scale. In addition there is a parameter that we call the entanglement rate, and denote by $\Gamma_i$, that is set by the time scale for quantum information and entanglement to spread from one end of the block to the other. Then $g_i = \Gamma_i / \Delta_i$.

Physically, $\tau_i = \Gamma_i^{-1}$ can be viewed as an “entanglement Thouless time” for block $i$ in the following sense: Put one end of block $i$ in strong local contact with a much longer block $j$ that is a good conductor $g_j \gg 1$. Initialize this two-block system in a pure product state with no entanglement between the two blocks. We choose typical random initial pure states of the two blocks, not eigenstates. Under the unitary time evolution of these two coupled (but otherwise isolated) blocks, the entanglement entropy will then grow and saturate on time scale $\tau_i = 1/\Gamma_i$, with the final value close to the thermal equilibrium entropy of the smaller block $i$. Thus we can call $\tau_i$ the entanglement time of block $i$; it is the time for entanglement to spread across the full length of the block. In principle, with knowledge of the microscopic couplings...
one could attempt to compute this time, but here the $\Gamma_i$ of the blocks are taken as inputs for the RG scheme.

It is noteworthy that $\tau$ is not the energy transport time. The latter, denoted by $\tau_{tr}$, is the time scale to relax an extensive energy imbalance across the block. On the end-to-end entanglement time-scale $\tau_i$ the amount of energy transported across the block remains of order the microscopic energy scale, so is not extensive. To relax an extensive energy imbalance requires transporting an extensive (in $l_i$) amount of energy, so requires of order $l_i$ entanglement times. Hence $\tau_{tr} \sim l_i \tau_i$. Note that the entanglement time $\tau_i$ is well defined even in a system subject to external periodically time dependent fields, such as a Floquet system, where total energy is not conserved and there is no extensive quantity that can be transported, so the transport time is meaningless.

The two-block parameters, $\Gamma_{ij}$, $\Delta_{ij}$ and $g_{ij} = \Gamma_{ij}/\Delta_{ij}$, are defined as the block parameters that would ensue if the two adjacent blocks are treated as a single block. For instance $\Delta_{ij} = W 2^{-l_i - l_j} \sqrt{l_i + l_j} \approx \Delta_i \Delta_j / W$. We call the link between these adjacent blocks $i$ and $j$ “effective” if $g_{ij} \gg 1$ and “ineffective” if $g_{ij} \ll 1$. A general requirement to be met by the initial distributions and retained throughout the RG flow is that the smallest block rate $\min_i \Gamma_i$ is larger than the largest two-block rate $\Omega = \max_{ij} \Gamma_{ij}$. $\Omega$, the largest two-block rate, serves as the running RG frequency cutoff scale. In this way all the fast rates ($\Gamma > \Omega$) are intra-block, while the slow rates, below the cutoff scale, are inter-block.

We now frame the RG as a strong disorder scheme operating on the chain in real space. At each RG step the cutoff scale $\Omega$ is reduced by joining the two blocks with the largest inter-block rate $\Gamma_{ij}$. Thus the old two-block parameters become the new one-block parameters of this new larger block. The non-trivial part of the renormalization is to determine the new two-block parameters $\Gamma_L$ and $\Gamma_R$, which connect the new block to its left and right neighbors. To compute these rates we have to solve for the entanglement rate of three coupled blocks. This calculation cannot be done microscopically in the most general case, but the structure of the solution is rather constrained by the known behavior in limiting cases. These constraints allow us to formulate a closed and self-consistent RG scheme. Modifying details of the RG scheme within the allowed constraints does not significantly change the outcome.

Suppose we are now joining blocks 1 and 2 with the fastest two-block rate $\Gamma_{12}$ and want to find the new rate $\Gamma_R$ of the three block system 1, 2, 3. There are two limits in which we can obtain simple reliable expressions for this rate. First, if both links are ineffective, $g_{12} \ll 1$ and $g_{23} \ll 1$, then we can compute $\Gamma_R$ by straightforward perturbation theory in the weak dimensionless couplings (see appendix A) to obtain

$$\Gamma_R = \Gamma_{12} \Gamma_{23} / \Gamma_2. \quad (1)$$

This case describes the process of making a bigger insulator out of two insulating links. When applied repeatedly to a long insulating chain this rule indeed leads to the expected exponential increase of the entanglement time with the length of the insulator.

Second, if both links lead to effective coupling, $g_{12} \gg 1$ and $g_{23} \gg 1$, then the entanglement spreads sequentially through the three block chain and we must add the entanglement times $G^{-1} = \Gamma^{-1}$:

$$\frac{1}{\Gamma_R} = \frac{1}{\Gamma_{12}} + \frac{1}{\Gamma_{23}} - \frac{1}{\Gamma_2}. \quad (2)$$

In a system with energy conservation the above formula is simply Ohm’s law for the thermal resistances.

The two RG rules given above lead to the correct scaling of length and time in insulating regions ($l \sim \log \tau$) and fully conducting regions ($\sim \tau \sim \tau_{tr}/l$). To complete the RG scheme we have to determine the behavior of boundaries between insulating and conducting regions. There we expect to encounter three block systems with one effective link $g_{12} \gg 1$ and one ineffective link $g_{23} < 1$. We have to distinguish the case in which the effective link is a link between two metallic blocks from the case when it is a link between a metallic and an insulating block.

Joining an insulator and a conductor, even if the link ultimately turns out to be effective, leads to exponential suppression of the relaxation rate with the length of the insulator (see appendix A). Coupling this structure to yet another insulator (i.e. the ineffective link $g_{23}$) would lead to further exponential suppression and hence insulating-like scaling of $\Gamma_R$ as prescribed in the RG rule (1). In appendix A we justify this formula using resummation of the perturbation theory. On the other hand, if the effective link $g_{12}$ is between two conductors, then the exponential suppression of the rate $\Gamma_R$ is only the result of the transport through the ineffective link $g_{23}$ which constitutes the bottleneck for entanglement spread. In this case we should use the second RG rule (2), which essentially adds the time of this bottleneck to the faster timescale $\Gamma_{12}^{-1}$.

We did not give expressions for intermediate regimes where $g_{ij} \sim 1$. Our approach will rely on having such a wide distribution of $g$’s at the interesting fixed point, that the probability of having $g \sim 1$ on a link vanishes. In practice we thus treat any $g > 1$ as $g \gg 1$, and any $g < 1$ as $g \ll 1$.

### III. FIXED POINTS AND FLOWS

Application of these RG rules to a chain with a random distribution of coupling constants leads to a flow of those distributions. Instead of solving the rather complicated integro-differential equations for the scale-dependent distributions we simply simulate the RG process on an ensemble of chains, each with up to $10^5$ or more initial blocks. Each block in the initial state is taken to be a $100 \times 100$ matrix with uniform $\Delta = W/100$ and $g = 1$, so the initial block lengths are $l_0 = \log_2(100)$. This immediately implies also a uniform $\Delta_{ij}$. The randomness is...
introduced in the distribution of the inter block couplings $g_{ij}$, which are generated in the following way. First a set $g_{ij}$ is drawn from a log normal distribution with mean $\langle \log(g_0) \rangle$ and standard deviation $\sigma_g = 1$. The problem with this initial set, however, is that the link entanglement times $\tilde{\tau}_{ij}$ obtained from these couplings do not necessarily satisfy the requirement that all link times must be longer than the individual block times $\tau_l = \tau_0$ (taken to be constant initially). To guarantee this hierarchy we adjust the link times by adding to them the adjacent block times $\tilde{\tau}_{ij} = \tilde{\tau}_{ij} + 2\tau_0$. The new dimensionless link couplings $g_{ij}$ are now obtained from the adjusted link times $g_{ij} = 1/(\tau_{ij}\Delta_{ij})$. We use the parameter $\langle \log(g_0) \rangle$ as the tuning parameter for the many-body localization transition. Although we start with moderate randomness, and only on the links, near the critical point the RG flows rapidly to strong randomness in all parameters.

The RG flow of the distribution of coupling constants near it.

A natural scaling variable to consider in our case, which reflects the link entanglement rates, is $\gamma = \ln(\Omega/\Gamma_{\text{link}})$. The flow of this variable is characteristic of an infinite randomness critical point. Specifically, we find a linear growth of the standard deviation $\delta\gamma$ with the RG flow parameter $\Lambda = \log(\Omega/\Omega_0)$, where $\Omega_0$ is the bare energy scale. This is the same as in the well known random singlet phase discussed by Fisher. However, as we will see below, the scaling between length and time scales at the critical point is different. Here we find log($t$) $\sim t^0$ with $\psi = 1$, compared to $\psi = 1/2$ at the random-singlet ground state infinite-randomness fixed point. Thus in this sense the flow to infinite randomness is stronger at our new fixed point than it is at the ground state infinite-randomness fixed points, and this distinction has important consequences.

The flow to infinite randomness is important also as a posteriori justification of the RG scheme. In the formulation of the RG steps we have assumed that the links realize extreme situations with either $g \ll 1$ or $g \gg 1$. As shown explicitly in appendix B3 the flow of the scaling variable $\gamma$ to infinite randomness is accompanied by broad distribution of the dimensionless coupling constant $g = \Gamma/\Delta$, so that near the critical point almost all links indeed realize extreme values of $g$. In the following sections we analyze the universal dynamics and the scaling of physical quantities on either side of the transition, where they are influenced by the infinite randomness critical point.
IV. ENERGY TRANSPORT AND ENTANGLEMENT SPREAD

An obvious property to study in systems with energy conservation is the behavior of thermal transport near the many-body localization transition. Information on the thermal transport can be gained directly from the RG flow by inspecting how the typical thermalization time of a block $\tau_{tr} = 1/T$ scales with the block size $l$. In the insulating phase we expect that $\tau_{tr}(l)$ grows exponentially with $l$, or $l \sim \log \tau_{tr}$, while in a diffusive conductor $l \sim \sqrt{D \tau_{tr}}$, where $D$ is the diffusion constant. One might expect that $D$ vanishes continuously as the transition is approached. However, this is not the result we find from the RG flow. Rather, we obtain a length-time scaling which follows a generalized power-law scaling $l \sim \tau_{tr}^{\alpha_c}$. As seen in Fig. 2(b), far from the critical point we indeed have the diffusive $\alpha = 1/2$, but closer to the transition the thermal transport is subdiffusive. The exponent $\alpha$ vanishes continuously at the critical point, where the length-time scaling becomes logarithmic as in the localized phase.

From the anomalous diffusion exponent $\alpha$ we can immediately infer the rate of entanglement entropy growth in a system undergoing time evolution from an initially nonentangled product pure state. The bipartite entanglement entropy across a link in our chain generated after time $\tau$ is proportional to the number of degrees of freedom that become entangled by that time, i.e. $S \sim l(\tau)$. Substituting $\tau_{tr} = lT$ into the relation $l \sim \tau_{tr}^{\alpha_c}$ we then find $S \sim \tau^{\alpha/(1-\alpha)}$. In particular this scaling relation implies ballistic entanglement spreading ($S \propto t$) in systems with diffusive energy transport, as already noted in Ref. [22]. On the other hand the two exponents $\alpha$ and $\alpha_{ent}$ have the same asymptotic behavior at the critical point as $\alpha$ and $\alpha_{ent} \rightarrow 0$.

V. EFFECTIVE GRIFFITHS PHASE MODEL

We argue that the existence of a subdiffusive phase is a natural precursor to the many-body localization transition in one dimension. If we assume that the many-body localization transition is continuous, then it must be accompanied by a diverging correlation length, $\xi \sim |g_0 - g_0c|^{-\nu}$. As we show below, we do indeed find such a power law divergence of the correlation length. If we look at the system at this length scale, then it looks critical. Since this is a critical point governed by an infinite randomness fixed point with $\psi = 1$, regions of this critical system viewed at this length scale $\xi$ show a wide range of local behavior, ranging from insulating to thermalizing, with blocks of length $\xi$ being either critical or insulating, each with a probability of order one. For a system that is globally delocalized ($g_0 > g_0c$), on longer scales $l$ than $\xi$ the system is typically locally thermalizing, but longer locally critical blocks of length $l$ may exist with a probability that behaves as $p(l) \sim \exp(-l/\xi)$. While they are exponentially rare, such long critical or insulating regions lead to an exponentially long delay of the entanglement time $\tau(l) \sim \tau_0 \exp(l/\xi_0)$, where $\xi_0$ and $\tau_0$ are microscopic length and time scales respectively. Hence these rare critical regions have a significant effect on the average; this is a defining feature of a Griffiths regime [23].

In a long section of length $L \gg \xi$, the typical length $l_m$ of the longest locally critical block is given by $p(l_m) \sim \xi/L$, which gives $l_m = \xi \log(L/\xi)$. Near enough to the critical point, these rare, long critical blocks are the dominant bottlenecks to entanglement spread and energy transport. Substituting $l_m$ in the exponential for the time scale we find $\tau \sim L^z$ and $\tau_{tr} \sim L^{z+1}$, with continuously variable Griffiths dynamical exponent $z = \alpha_{ent}^{-1} \approx \xi/\xi_0$. Note that if we instead had $\psi < 1$, as at the ground state infinite randomness critical points, then $\tau(l) \sim \exp(\psi l)$, which would be too weak to produce this subdiffusive behavior, so the dominant Griffiths domains are not critical regions, but insulating regions.

These Griffiths effects dominate the long time transport as long as $z > 1$, while the systems has “normal” transport farther from the transition, where $z$ “sticks” to the value $z = 1$ that gives ballistic entanglement spread and diffusive energy transport. Note this also implies that the Griffiths exponent for entanglement spreading behaves as $\alpha_{ent} \sim (g_0 - g_0c)^{\psi}$ as one approaches the transition. Fig. 7 of Appendix C shows that our results are consistent with this behavior. Near the transition, these Griffiths effects lead to a broad distribution of the dimensionless coupling $g(L)$ at large $L$, due to the variation in the severity of the slowest bottleneck. This then matches on nicely to the broad distributions we find at the critical point.

VI. SCALING OF EIGENSTATE ENTANGLEMENT ENTROPY

So far we considered only the properties of a system undergoing time evolution. From this perspective the critical point is similar to the many-body localized phase in that both exhibit transport or entanglement times that grow exponentially with the length. However, more direct insight on the nature of the critical point can be gleaned from studying how generic energy eigenstates change across the critical point. As mentioned in the introduction, the many-body localization transition represents a novel type of critical point at which the eigenstate entanglement scaling changes from area-law to volume law [10, 24, 25]. The real space RG approach can lend information on how this change takes place.

First, we explain the relation between the dimensionless coupling $g$ and the entanglement entropy in eigenstates. Suppose we renormalized the chain all the way down to the point where we have only two blocks remaining in the system. If these two blocks were decoupled then the exact eigenstates would be non-entangled
link somewhere in the chain. As a toy example consider a chain of three blocks, where blocks 1 and 2 are coupled and together span the interface, whereas blocks 2 and 3 are completely disconnected (i.e. $\Gamma_{23} = 0$). In this case we will first join blocks 1 and 2 to get a new block with $g_{12} > 0$, which spans the interface. Obviously there is entanglement across the interface, which $S_{12} = \log(1 + g_{12})$ represents. However if we now continue to renormalize we would obtain $g = 0$ for the last remaining block, which of course represents only the absence of entanglement across the disconnected link.

The RG scheme is repeated on a large number of disorder realizations allowing to obtain a full distribution of the entanglement entropy. Examples of entropy distributions found in the different states, including the localized state, the critical point, the Griffiths phase and the diffusive regime are shown in appendix D. Here in Fig. 3(a) we present the average and standard deviation of the entropy as a function of the bare coupling $\langle \log g_0 \rangle$ calculated for varying system sizes $L$ (in units of elementary blocks of $l_0$ spins). The entropy and its fluctuation are normalized by the extensive thermal entropy $S_T = L l_0 \log 2$. As expected, the variation of $S/S_T$ and $\delta S/S_T$ as a function of $\langle \log g_0 \rangle$ sharpen with increasing size in a way which suggests the existence of a critical point in the limit $L \to \infty$. In this case we anticipate that near the critical point the functions $S(g_0, L)/S_T$ and $\delta S(g_0, L)/S_T$ should all collapse on scaling functions of a single variable $(g_0 - g_0_c) L^{1/\nu}$, where the critical value $g_0_c$ and the universal critical exponent $\nu$ are fitting parameters (see Fig. 3(b)).

The correlation-length critical exponent extracted from the data collapse, $\nu \approx 2.8$, satisfies the Harris inequality $\nu \geq 2/d$ required for stability of the critical point 20,21. It is interesting that a much smaller exponent, which violates this inequality, was found in recent finite size scaling analysis of exact diagonalization data 16,28. This, as well as other differences from our scaling form may be due to the small system sizes studied in Refs. 16,28, $L < 18$, which are likely to be too small to approach the scaling limit. Indeed in our case, although we start from a coarse grained model, system sizes of 50 or more blocks are needed. We do make some assumptions in choosing some of the details of our RG rules. There thus remains the possibility that the precise value of this exponent $\nu$ is sensitive to such details, so this estimate of $\nu$ should be viewed as possibly approximate.

The entanglement scaling functions shown in Fig. 3(b) expose important properties of the subdiffusive Griffiths phase. A finite size system in the subdiffusive Griffiths phase corresponds to a point on the positive scaled $x$ axis. With increasing system size, such a system flows on the scaling functions to the right hand side of the graph, where the entropy density approaches its thermal-equilibrium value. At the same time the relative fluctuations of the entropy in the Griffiths phase vanish with system size. These results indicate that the entire Griffiths regime is fully thermal in the limit of large $L$.
More information on the critical point itself can be gleaned from inspecting the full distribution of the entropy. The result shown in appendix Fig. 3(b) suggest that the entropy distribution approaches a power-law \( p(s) \sim 1/s^\zeta \) with \( \zeta \approx 0.84 \). For \( \zeta < 1 \) the average and standard deviation of the entropy density is expected to be a non vanishing constant. The fact that \( \zeta \) is close to 1 may explain why these constants appear to be very small. We note that if \( \zeta = 1 \) then the entropy density at the critical point would approach zero as \( 1/\log L \) and its fluctuations would vanish as \( 1/\sqrt{\log L} \).

**VII. CONCLUSIONS**

We presented a new renormalization group framework, which provides a description of the many-body localization transition in one-dimensional systems. The dynamical scaling between length and time \( t \sim l^\nu \) within the thermal phase extracted from the RG calculation shows that the transition occurs at a critical point where \( \alpha \) vanishes continuously. Hence the delocalized phase near the critical point displays sub-diffusive transport and sub-ballistic entanglement growth in time evolution. This behavior is understood in terms of a Griffiths phase dominated by rare critical inclusions in a conductor.

We pointed out a connection between dynamical properties and the entanglement entropy associated with individual eigenstates near the critical point. Using this relation we show how eigenstates with area law entanglement in the localized phase transition to ones with volume law entanglement entropy characteristic of thermal states. This occurs through an infinite randomness critical point at which the distribution of entanglement entropy becomes broad, spanning the entire range from \( S \sim 0 \) to the full thermal value \( S \sim L \). The RG flow to infinite randomness is as strong as is possible, with exponent \( \psi = 1 \). In the delocalized Griffiths phase, on the other hand, the entanglement entropy density is peaked near the thermal value with fluctuations that vanish in the limit of large \( L \), indicating that this delocalized Griffiths regime is fully thermal in spite of its anomalous transport properties. More generally, the variation of the entropy and its fluctuations across the transition is expressed in terms of finite-size scaling functions from which we extract an estimate of the critical exponent \( \nu \equiv 2.8 \) associated with the diverging correlation length.

It is interesting to understand our result in view of the constraints set on the many-body localization transitions by the strong sub-additivity property of the entanglement entropy\[25\]. The constraint relevant to the situation we describe is that the critical point marking a direct transition to a thermal delocalized state must itself obey the eigenstate thermalization hypothesis and show thermal behavior of the entanglement entropy. At first sight this may appear to contradict our finding of strongly fluctuating, non thermal entanglement entropy at the critical point. However we note that strong subadditivity requires only thermal behavior of the the entanglement entropy associated with a subsystem of size \( l \) much smaller than the full system size \( L \). That is \( S(l, L) \) must behave as the thermal entropy with vanishing fluctuations in the appropriate thermodynamic limit \( L \rightarrow \infty \) while \( l/L \rightarrow 0 \). Thus we conclude that the critical point represents a weaker class of thermal states than the delocalized Griffiths phase. In the latter the entanglement entropy density for half of the system is thermal with vanishing relative fluctuations in the \( L \rightarrow \infty \) limit, while for the former only the limit of small subsystems \( (l \ll L) \) is fully thermal.

**ACKNOWLEDGMENTS**

Illuminating discussions with Anatoli Polkovnikov, Gil Refael, Dmitri Abanin, Rahul Nandkishore, Sarang Gopalakrishnan, and Joel Moore are gratefully acknowledged. EA thanks the Miller Institute at UC Berkeley, the Aspen Center for Physics under NSF Grant # 1066293 and the Perimeter Institute for hospitality. EA and RV were supported by ERC grant UQUAM, the ISF grant # 1594/11 (EA).

**Appendix A: Derivation of the RG rules**

In this section we derive the RG rules, which are used in the main text. At each step of the RG we join the pair of blocks connected by the fastest link rate \( \Gamma_{i,i+1} \) into a single block thereby making the link variables of this pair into the new block variables. The non trivial part of the transformation prescribes what are the new link rates \( \Gamma_L \) and \( \Gamma_R \) connecting the newly joined block to its left and right neighbors. Before explaining how to compute these rates we shall discuss the physical meaning of the input two-block rates.

1. Two block relaxation

The “bare” two-block relaxation rates \( \Gamma_{ij} \) are given as input and not directly calculated. However we need to know how they depend on the microscopic coupling matrix elements between blocks in order to understand how these rates enter the calculated three-block rates. In general we want to consider a situation of two neighboring blocks 1 and 2 characterized by internal rates \( \Gamma_1 \) and \( \Gamma_2 \) and level spacings \( \Delta_1 \) and \( \Delta_2 \). Recall that these blocks are really chains of microscopic constituents, e.g. spins. Therefore, for a Hamiltonian system with energy conservation the band-width that captures almost all of the many body spectrum grows with the block length as \( W\sqrt{l} \), where \( W \) is a microscopic energy scale and \( l \) is the block length. Correspondingly the typical many-body level spacing for the block is \( \Delta \sim W\sqrt{l}/2^l \). In
a Floquet system on the other hand the bandwidth remains constant $W$ and therefore the mean level spacing $\Delta \sim W/2^l$. In practice this difference is sub-leading to the exponential dependence and will make no difference for the critical point.

In absence of coupling between the blocks the eigenstates of the two block system are, of course, products of the single block eigenstates. We now introduce coupling between the two blocks through a local operator $\hat{J}_{12}$ which changes microscopic degrees of freedom on the two neighboring edges of the blocks. It can be written as

$$\hat{J}_{12} = J_{12} \left( A_1^\dagger A_2 + \text{h.c.} \right),$$

(A1)

where $A_1$ operates on the edge of block 1 and $A_2$ operates on the edge of block 2.

If the system is prepared in a product of the single block eigenstates, the coupling leads to decay of the state by inducing transitions to other product states. The Fermi golden rule expression for this decay rate is

$$\Gamma_{12} \sim |J_{12}|^2 \sum_{n_1, n_2} \left| \langle i_1 | \hat{A}_1 | n_1 \rangle \right|^2 \left| \langle i_2 | \hat{A}_2 | n_2 \rangle \right|^2 \delta (\omega_1 + \omega_2)$$

where $|i_1, i_2\rangle$ is the initial state and the summation is over the possible final states $|n_1, n_2\rangle$. Here, $\omega_1 = E_{n_1} - E_{1}$ and $\omega_2 = E_{n_2} - E_{2}$ are the energy changes due to the transitions in block 1 and 2 respectively.

The nature of the transition matrix elements $\langle n_b | A_1 | m_a \rangle$, relevant for relaxation from one side of the block to the other, depend on whether the block $b$ under consideration is delocalized or localized. If it is a strongly localized block, we can write the block eigenstates as mutual eigenstates of quasi-local integrals of motion $\tau_i^{z_l}$ (“l-bits”), $|n\rangle = |\tau_{1},\ldots,\tau_{l}\rangle$. The local operator $A_1^\dagger$ at the edge of the block can be written in terms of these l-bits as

$$A_1^\dagger = Z^{\tau_1^{z_l}} + \sum_{r=1}^{l} e^{-r/\xi_0} \sum_{n} a_{n,r} \hat{O}_{nr}$$

(A3)

where the operators $\hat{O}_{nr}$ are non local operators that flip multiple l-bits extending up to a distance $r$ from the first site. $a_{n,r}$ are random coefficients of order 1 and random sign and $\xi_0$ is a microscopic length scale ($\xi_0 \leq 1$). Since we are interested in the end to end relaxation we only consider transitions, which change the state of the integrals of motion $\tau_i$ all the way to the other side of the block. For typical matrix elements of interest we have: $\langle n | A_1 | m \rangle = e^{-l/\xi_0} a_{nm}$, where $a_{nm}$ are random numbers of order 1 drawn from a state independent distribution as long as the two states $|m\rangle$ and $|n\rangle$ differ by an energy of up to order $W$. $a_{nm}$ essentially vanish for larger transition energies (e.g. for transitions of the order of the bandwidth $W/\sqrt{l}$).

On the other hand, when dealing with a delocalized block the single block integrals of motion are the projectors on single block eigenstates $|n\rangle \langle n|$, which are highly non local operators. We take the transition matrix elements of the local operator $A_1$ between these states to be functions of the energy difference between them alone:

$$\left| \langle n | A_1 | m \rangle \right|^2 = F_b(\omega_{nm}).$$

(A4)

These matrix elements are directly related to the temporal decay of the autocorrelation function $f_b(t) = \langle n | A_1^\dagger(t) A_1(0) | n \rangle$ through a Fourier transform

$$F_b(\omega) = \frac{1}{\rho_b} \int dt e^{-i\omega t} f_b(t),$$

(A5)

where $\rho_b \sim 1/\Delta_b$ is the density of states of block $b$. For example if the block is diffusive and the system is energy conserving then $f_b(t) \sim \sqrt{\rho_0 t}$. Note that we can unify the notations for the different cases in the following side we use $F_b(\omega) = e^{-2l/l_0/\xi_0}(W - |\omega|)$.

We are now ready to evaluate the relaxation rates by converting the sums in (A2) into integrals over the respective density of states and plugging in the appropriate matrix elements.

$$(A2) \Gamma_{12} = J_{12}^2 \rho_2 \int_{-W}^{W} d\omega F_1(\omega) F_2(-\omega) = J_{12}^2 \int_{-\Gamma_1}^{\Gamma_2} dt f_1(t) f_2(t)$$

(A6)

The upper cutoff is set by the minimum of the decay times of the two blocks, $\tau = \min(\Gamma_1^{-1}, \Gamma_2^{-1})$.

Let us pause to consider the relaxation rate in different cases, i.e. when we couple (i) two insulators, (ii) two conductors or (iii) an insulator and a conductor. In case (i) taking for simplicity $\xi_1 = \xi_2 = \xi_0$ and $J_{12} \approx W$, the microscopic energy scale, we have:

$$\Gamma_{12} = \frac{W^3}{\Delta_1 \Delta_2} e^{-l/(12l)/\xi_0}$$

(A7)

Now, using $\Delta_b = W^{4/9}/2^{1/2}b$ and $\Delta_{12} = W^{4/21}/2^{9/12}$ we can express the dimensionless coupling as

$$g_{12} = \frac{1}{\sqrt{l_{12}^{1/12}} \exp[(\ln 2 - \xi_0^{-1})/12]} \ll 1$$

(A8)

This must be smaller than 1 because a similar expression, with the same exponential factor, holds for $g_1$ and $g_2$ of the individual blocks, and for the latter to be much smaller than 1, as assumed, we must have $\ln 2 < \xi_0^{-1}$.

In case (ii), when both blocks are conducting, perturbation theory is not valid, but at least it can indicate that the coupling must be effective and the blocks thermalize. Substituting $f_0(t) = 1/\sqrt{Wt}$ in (A6) we obtain

$$\Gamma_{12} \approx W \log(W\tau_{b0}) \gg \Delta_{12}.$$  

(A9)

In this case we expect the end-to-end entanglement time is simply the sum of the times to entangle across each block: $\Gamma_{12}^{-1} = \Gamma_1^{-1} + \Gamma_2^{-1}$.

Finally in case (iii) of a conductor coupled to an insulator we find $\Gamma_{12} = \Gamma_1^{-1} + (W^2/\Delta_2)e^{-2l/l_2}$. In the case of
a fast conductor ($\Gamma_1 \gg \Gamma_2$) we have for the dimensionless coupling:

$$g_{12} \sim e^{2i_{2}(\ln 2-\xi_2^{-1})}e^{1.1n_{2}}$$  \hspace{1cm} (A10)

If the length $l_2$ of the insulator is long enough, the incipient conductor of length $l_1$ is not able to thermalize it.

2. Two block entanglement rate

If the two block relaxation rate calculated above turns out to be smaller than the two block level spacing $\Delta_{12}$ then $g_{12} < 1$, it is deemed ineffective and the two blocks do not exhibit end-to-end relaxation. However, there is still a physical rate, which describes the rate at which the degrees of freedom at the furthest ends of the blocks get entangled with each other. We will see that the expression for the entanglement rate of two blocks that end up insulating turns out to be identical to the expression [A6] for the relaxation rate.

The fact that the coupling between the blocks is ineffective means that operators, which were integrals of motion of the individual blocks map continuously to integrals of motion of the two block system. In particular the projectors on single block eigenstates $|n_1\rangle \langle n_1| \otimes 1$ and $1 \otimes |n_2\rangle \langle n_2|$ are continuously connected to integrals of motion of the coupled two block system. The coupling between the two blocks generates a diagonal interaction between these conserved quantities, which can lead to generation of entanglement in the course of time evolution. We want to find the effective diagonal coupling generated between degrees of freedom that, if localized, are located at opposite ends of the two block system.

The diagonal interaction we are interested in is generated by second order of perturbation theory in the local (off diagonal) inter-block coupling:

$$V(n_1, n_2) = J_{12}^2 \sum_{m_1, m_2} |\langle n_1 | A_1 | m_1 \rangle|^2 |\langle n_2 | A_2 | m_2 \rangle|^2 = J_{12}^2 \rho_1 \rho_2 \int d\omega_1 d\omega_2 \frac{F_1(\omega_1)F_2(\omega_2)}{\omega_1 + \omega_2 + i\eta}$$

$$= J_{12}^2 \int dt_1 dt_2 d\omega_1 d\omega_2 e^{-i\omega_1 t_1 - i\omega_2 t_2} \frac{f_1(t_1)f_2(t_2)}{\omega_1 + \omega_2 + i\eta} = J_{12}^2 \int_0^\infty dt f_1(t)f_2(t)$$  \hspace{1cm} (A11)

This is exactly the same expression we obtained for the relaxation rate. It is important to note here that matrix elements $\langle m_b | A_b | n_b \rangle$, which lead to generation of end to end interaction (and through it end to end entanglement) are only those which involve end-to-end couplings between degrees of freedom at the far ends of the two block system. Hence, as in the previous section, we are interested in the non-local tail of the operator $A_b$, when written terms of the block integrals of motion [A3]. For this reason the function $F_b(\omega)$ in an insulating block involves a suppression factor of order $e^{-2\xi_b/\xi_3}$

3. Perturbative three block relaxation

Suppose we are now joining blocks 1 and 2 with the fastest link rate $\Gamma_{12}$. We must then find the new rate $\Gamma_{R}$ needed for thermalization (or end-to-end entanglement in the insulating case) through the three block system 1,2,3. We will see that this rate can be expressed in terms of the two-block and single block rates.

The simplest case to treat is when both of the links are ineffective, i.e. $g_{12} \ll 1$ and $g_{23} \ll 1$. In this case the decay rate from initial state $|i\rangle$ to final state $|f\rangle$ is obtained using the generalized Fermi-golden rule

$$\Gamma = 2\pi \sum_f |\langle f | T | i \rangle|^2 \delta(E_f - E_i).$$  \hspace{1cm} (A12)

with the $T$-matrix given by

$$T = \frac{1}{E_i - H_0 + i\eta} \frac{1}{E_i - H_0 + i\eta} \frac{1}{E_i - H_0 + i\eta} = \frac{1}{E_i - H_0 + i\eta} T,$$  \hspace{1cm} (A13)

where $H_0$ is the Hamiltonian of decoupled blocks (i.e. contains only the intra-block interactions) and $\hat{J}$ is the coupling between the blocks. In our case, $\hat{J} = \hat{J}_{12} + \hat{J}_{23}$. Clearly, to lowest order in $\hat{J}$ we recover the usual fermi golden-rule.

A crucial point is that the relaxation process we calculate involves a decay from an initial state $|i_1, i_2, i_3\rangle$ to a final state $|f_1, f_2, f_3\rangle$, which is different from the initial state in at least the first and the last labels (i.e $i_1 \neq f_1$ and $i_3 \neq f_3$). Otherwise it would not correspond to full end-to-end relaxation of the three block system.

The matrix elements of the $T$-matrix now take the explicit form
\( \langle f \, | \, T \, | \, i \rangle = \langle f \, | \, \hat{J} + \frac{1}{E_i - H_0 + i\eta} \hat{J} + \ldots \, | \, i \rangle \)
\[= \langle f \, | \, (\hat{J}_{12} + \hat{J}_{23}) \, | \, i \rangle + \langle f \, | \, (\hat{J}_{12} + \hat{J}_{23}) \frac{1}{E_i - H_0 + i\eta} (\hat{J}_{12} + \hat{J}_{23}) \, | \, i \rangle + \ldots \]
\[= \sum_m \langle f \, | \, (\hat{J}_{12} + \hat{J}_{23}) \, | \, m \rangle \frac{1}{E_i - E_m + i\eta} \langle m \, | \, (\hat{J}_{12} + \hat{J}_{23}) \, | \, i \rangle + \ldots \] \hspace{1cm} (A14)

The first order term dropped because \( | \, i \rangle \) and \( | \, f \rangle \) are not connected by a single application of \( \hat{J}_{12} \) or \( \hat{J}_{23} \). We also introduce a sum over complete intermediate states \( | \, m \rangle = | \, m_1, m_2, m_3 \rangle \). At this point the summation of \( m \) runs over the combinations of all the states of all the blocks. However, since eventually the intermediate states enter matrix elements of the form \( \langle f \, | \, \hat{J}_{23} \, | \, m \rangle \langle m \, | \, \hat{J}_{12} \, | \, i \rangle \), most of the combinations give zero contribution. Only when \( m_1 = f_1 \) and \( m_3 = f_3 \) the matrix element is non-zero. We are left with one summation over the label \( m_2 \) associated with the energy levels of the middle block.

For the summation over intermediate states of the second block we choose a basis of block eigenstates that is not the eigenstate basis. Rather, to take advantage of our knowledge of the intra-block rate \( \Gamma_2 \) we divide the middle block to two halves and take a basis of product states of the two halves. These states are broadened by \( \Gamma_2 \) (by the definition of this internal block rate), that is their energy can be effectively taken to be \( E_m + i\Gamma_2 \). In particular, this will give rise to an imaginary part of the energy denominator \( \eta = \Gamma_2 \) when evaluating the \( T \)-matrix element (A14).

We are now in position to compute the decay rate using (A12) and (A14) to obtain

\[ \Gamma = \sum_{f, m_2} \frac{\langle f \, | \, \hat{J}_{23} \, | \, m \rangle^2 \langle m \, | \, \hat{J}_{12} \, | \, i \rangle^2}{(E_i - E_m)^2 + \Gamma_2^2} \delta(E_f - E_i). \] \hspace{1cm} (A15)

Plugging (A1), defining the energy shifts within the blocks \( \omega_1 = E_{f_1} - E_{i_1}, \omega_{2L} = E_{m_2} - E_{i_2}, \omega_{2R} = E_{f_3} - E_{i_3}, \) and converting the sum into an integral over the density of states we get

\[ \Gamma = J_{12}^2 J_{23}^2 \int d\omega_1 d\omega_2L d\omega_2R d\omega_3 \frac{F_1(\omega_1) F_2L(\omega_{2L} + i\Gamma_2) F_2R(\omega_{2R} + i\Gamma_2) F_3(\omega_3)}{(\omega_1 + \omega_{2L} + \omega_{2R} + \omega_3) (\omega_1 + \omega_{2L})^2 + \Gamma_2^2} \delta(\omega_1 + \omega_{2L} + \omega_{2R} + \omega_3) \]
\[= J_{12}^2 J_{23}^2 \int d\omega d\epsilon_L d\epsilon_R \frac{F_1(\epsilon_L + \frac{\omega}{2}) F_2L(-\epsilon_L + \frac{\omega}{2}, \Gamma_2) F_2R(-\epsilon_R - \frac{\omega}{2} + i\Gamma_2) F_3(\epsilon_R - \frac{\omega}{2})}{\omega^2 + \Gamma_2^2} \] \hspace{1cm} (A16)

In the second line we changed variables to \( \omega = \omega_1 + \omega_{2L}, \omega' = \omega_3 + \omega_{2R}, \epsilon_L = (\omega_1 - \omega_{2L})/2 \) and \( \omega_R = (\omega_3 - \omega_{2R})/2 \), and integrated over \( \omega' \). We now transform the integral over \( \epsilon_L \) and \( \epsilon_R \) to integrals over time

\[ \Gamma = J_{12}^2 J_{23}^2 \int dt \omega dtL dtR e^{i\omega(tL - tR)} e^{-\Gamma_2 |tL| + \epsilon_L} f_1(t_L) f_2L(t_L) f_2R(t_R) f_3(t_R) \]
\[= \frac{J_{12}^2 J_{23}^2}{\Gamma_2} \int dtL dtR e^{-\Gamma_2 |tL| - \epsilon_L} f_1(t_L) f_2L(t_L) f_2R(t_R) f_3(t_R) \] \hspace{1cm} (A17)

where in the second step we integrated over \( \omega \) using the Fourier transform of a Lorentzian. Now, because there is an independent cutoff \( \Gamma_2^{-1} \) on each of the integration times, set by the second exponential factor above, we can drop the cutoff on the time difference set by the first exponential. Hence up to multiplicative constants that would be irrelevant in the RG for strong disorder we have

\[ \Gamma \approx \frac{1}{\Gamma_2} \left[ J_{12}^2 \int dtL e^{-\Gamma_2 |tL|} f_1(t_L) f_2L(t_L) \right] \cdot \left[ J_{23}^2 \int dtR e^{-\Gamma_2 |tR|} f_3(t_R) f_2R(t_R) \right] \approx \frac{\Gamma_{12} \Gamma_{23}}{\Gamma_2} \] \hspace{1cm} (A18)
In the last step we used the result \([A6]\) for the two block relaxation.

We now consider the generalization of the result \([A18]\) to the case where one link is effective \(g_{12} \gg 1\) and the other is ineffective \(g_{23} \ll 1\). Strictly speaking each two-block rate entering the formula \([A18]\) represents the true two block rate only when that link is ineffective. However if the link is between a metal and an insulator (or two insulators) then the perturbative rate \([A6]\) nevertheless scales correctly with the length of the insulator. Formally, to obtain the quantitatively correct rate we can take the perturbation series to infinite order in the coupling \(J_{12}\) which amounts to replacing the bare matrix element with the renormalized \(T\)-matrix element \(T_{12}\) in \([A18]\) above.

The above generalized Fermi golden rule calculation of a three block rate is in principle valid as long as the outcome rate is much larger than the three block level spacing. Otherwise there are no irreversible transitions. Nonetheless, we expect the same result to hold based on the known behavior in the localized state. Indeed the three block system represents, in this case, a long localized insulating region, for which we know that the length time scaling is logarithmic. Taking the log of the RG rule \(\Gamma = \Gamma_{12} \Gamma_{23} / \Gamma_{2}\) we have \(\log \tau = \log \tau_{12} + \log \tau_{23} - \log \tau_{2}\). Now multiplying by the bare localization length which characterizes the length time scaling in the insulator \(l = l_0 \log \tau\) we see that the RG rule is equivalent in this regime to adding of the lengths of the three blocks: \(l_{\text{tot}} = l_{12} + l_{23} - l_{2}\).

[I am not sure the next paragraph is needed. EA]. One can also attempt to derive the entanglement rate directly in this case by computing the effective diagonal interactions (i.e. energy shifts) generated by the off-diagonal interactions between the blocks as we did for the two block system. However, in the three block system, the interactions which produce end to end entanglement are generated only in fourth order of perturbation theory and are more complicated to compute than the generalized Fermi golden rule rates. A simpler way to obtain the three block entanglement rate is to connect the chain of three blocks to two auxiliary thermalizing blocks to the left and right with rates \(\Gamma_L, \Gamma_R\). This is done in section \([A17]\) by looking at how the entropy and its variance change across the transition; a finite size scaling analysis was used to collapse the data and extract the location of the transition and the correlation length critical exponent. We verify this result here by applying a finite size scaling analysis to a different quantity extracted from the numerical results: the derivative \(d \langle \log(g) \rangle / dl\). The variation of this quantity across the transition, from a negative slope in the insulator and at the critical point to the maximal positive slope in the metal, is shown in Fig. 4(a). The crossover functions obtained for different system sizes are then collapsed in Fig. 4(b) using the same parameters \(g_0, \nu\) obtained by scaling of the entropy curves in the main text.

Having located the critical point we can turn to characterize the critical flow. Infinite randomness critical points.

Appendix B: The critical flow – Infinite randomness scaling

In this appendix we characterize the flow of distributions and coupling constants at the critical fixed point and demonstrate that it is indeed an infinite randomness fixed point. The first step is to accurately locate the fixed point. This was done in section \([A17]\) by looking at how the entropy and its variance change across the transition; a finite size scaling analysis was used to collapse the data and extract the location of the transition and the correlation length critical exponent. We verify this result here by applying a finite size scaling analysis to a different quantity extracted from the numerical results: the derivative \(d \langle \log(g) \rangle / dl\). The variation of this quantity across the transition, from a negative slope in the insulator and at the critical point to the maximal positive slope in the metal, is shown in Fig. 4(a). The crossover functions obtained for different system sizes are then collapsed in Fig. 4(b) using the same parameters \(g_0, \nu\) obtained by scaling of the entropy curves in the main text.

Having located the critical point we can turn to characterize the critical flow. Infinite randomness critical points.
are usually characterized by (i) how the scaling variables (associated with coupling constants) and their standard deviations scale with the block length \( l \), and (ii) how the block length scales with the logarithm of the energy cut-off \( \lambda = \log(\Omega_0/\Omega) \).

Fig. 3 clearly shows a linear dependence of the scaling variable \( \gamma = \log(\Omega/T_{\text{link}}) \), associated with the link entanglement rates, and its standard deviation \( \delta \gamma \), with the block length \( l \). The same linear scaling behavior is seen for \( \delta \log(g) \). The length time scaling at the critical point is shown in Fig. 3(b), showing a linear dependence of \( l \) on \( \lambda \), that is, a logarithmic dependence on the time.

These results indicate a different class of infinite randomness critical point from the famous cases of the random singlet phase and the random transverse field Ising model considered by Fisher [21, 29]. Specifically, at Fisher’s fixed point the RG flow parameter \( \lambda \) scales with \( l \) as \( \lambda \sim l^\psi \) with \( \psi = 1/2 \), while in our case \( \psi = 1 \).

**Appendix C: Length time scaling**

In Fig. 6 we present the length vs. time scaling extracted from the RG calculation. The lines show the average block length \( l \) versus the time-scale given by the inverse of the cutoff frequency \( t = \Omega^{-1} \). As explained in the main text this scale corresponds to the time for entanglement spreading. Therefore the power-laws fitted in the delocalized phase give the dynamical exponent for entanglement growth \( \alpha_{\text{ent}} \). To obtain the transport exponent \( \alpha \) shown in Fig. 2(b) of the main text we use the scaling relation derived in the main text \( \alpha = \alpha_{\text{ent}}/(1+\alpha_{\text{ent}}) \). We note that at the critical point and the localized phase, i.e. for \( \log(g_0) \leq \log(g_{0c}) \approx -1.8 \), the exponent vanishes and the dependence becomes instead logarithmic growth of entanglement with time.

We can ask how the dynamical exponent \( \alpha \) or \( \alpha_{\text{ent}} \) vanishes as \( g_0 \) approaches \( g_{0c} \). The plot of \( \alpha_{\text{ent}} \) versus \( (g_0 - g_{0c}) \) on a log-log plot, computed using the RG on systems of varying sizes is shown in Fig. 7. These results seem consistent with the expected \( \alpha_{\text{ent}} \sim (g_0 - g_{0c})^\nu \). Note however that it is rather hard to obtain the asymptotic behavior of \( \alpha \) near the critical point because of two requirements that have to be met. First, we must get sufficiently close to the critical point in order to be in the critical scaling regime. Second, the sub-diffusive transport is a property of the Griffiths phase, thus for given value of the tuning parameter \( g_0 - g_{0c} \) in the scaling regime we must obtain \( \alpha \) from systems sizes that are much larger than the long correlation length \( \xi = c(g_0 - g_{0c})^{-\nu} \). The quality of the power-law fit suggests that the systems we are calculating may be just barely reaching the asymptotic scaling of the exponent \( \alpha \).

**Appendix D: Entanglement entropy distributions**

In this appendix we show examples of the entanglement entropy distributions computed using the RG flow applied to an ensemble of disorder realizations. Fig. 8 shows four distributions taken respectively from the localized phase, the critical point, the Griffiths phase and the diffusive regime for long chains with \( L/l_0 = 10000 \). In the localized phase the entanglement entropy follows an area law, therefore the distribution of the specific entropy \( s = S/S_T \) is concentrated near zero, with the tail of the distribution consistent with a simple exponential. At the critical point the entanglement entropy shows a broad distribution that is consistent with a power law \( P_e(s) \sim 1/s^\zeta \) with \( \zeta \approx 0.8 \). In the Griffiths phase the distribution has a relatively narrow peak near the ther-
FIG. 8. (Distributions of the entanglement entropy divided by its thermal value for systems of length $L/l_0 = 10000$ calculated in the different regimes: the localized phase (top left), critical point (top right), Griffiths phase (bottom left), and diffusive regime (bottom right).

...mal value. Finally, in the diffusive phase the distribution becomes essentially a delta function at the thermal value minus a tiny finite-size correction.

[1] P. W. Anderson, Phys. Rev. 109, 1492 (1958).
[2] R. Nandkishore and D. A. Huse, ArXiv e-prints (2014), arXiv:1404.0686 [cond-mat.stat-mech].
[3] E. Altman and R. Vosk, ArXiv e-prints (2014), arXiv:1408.2834 [cond-mat.dis-nn].
[4] R. Vosk and E. Altman, Phys. Rev. Lett. 110, 067204 (2013).
[5] M. Serbyn, Z. Papić, and D. A. Abanin, Phys. Rev. Lett. 111, 127201 (2013).
[6] D. A. Huse and V. Oganesyan, Phys. Rev. B 90, 174202 (2014).
[7] J. Z. Imbrie, ArXiv e-prints (2014), arXiv:1403.7837 [math-ph].
[8] Y. Bahri, R. Vosk, E. Altman, and A. Vishwanath, ArXiv e-prints (2013), arXiv:1307.4092 [cond-mat.dis-nn].
[9] M. Serbyn, M. Knap, S. Gopalakrishnan, Z. Papić, N. Y. Yao, C. R. Laumann, D. A. Abanin, M. D. Lukin, and E. A. Demler, Phys. Rev. Lett. 113, 147204 (2014).
[10] M. Žnidarič, T. Prosen, and P. Prelovšek, Phys. Rev. B 77, 064426 (2008).
[11] J. H. Bardarson, F. Pollmann, and J. E. Moore, Phys. Rev. Lett. 109, 017202 (2012).
[12] D. J. Luitz, N. Laflorencie, and F. Alet, ArXiv e-prints (2014), arXiv:1411.0660 [cond-mat.dis-nn].