Princeton University

Department of Physics

submitted in partial fulfillment
of the requirement for the
degree of bachelor of arts

OPERATOR SPREADING AND ENTANGLING IN A QUANTUM CHAOTIC SPIN CHAIN

Author: Cheryne Jonay

Adviser: Professor David A. Huse

Second Reader: Professor Shivaji L. Sondhi

Abstract

We study operator dynamics and entanglement in a quantum chaotic Heisenberg spin chain. Specifically, we are interested in how these two processes - spreading and entangling - are related. What we find is that they are distinct processes: the dynamics of (i) the spreading of local operators, (ii) the width of the 'front' of this moving operator, (iii) the entanglement production, are characterized by different time scales. In a finite size system an initially local operator spreads from one end to the other on one time scale (given by the so-called butterfly speed v_B), but becomes an essentially random operator (a fully entangled operator) on the full system on a longer time scale (given by the slower entanglement speed v_{opEE}). The moving front of the operator broadens with time, but sublinearly, perhaps diffusively as $\sim \sqrt{t}$. For the spreading operator, we relate the local rate of entanglement production $\partial S/\partial t$ and the spatial gradient of the entanglement across the bipartite cut as the cut is moved, $\partial S/\partial x$. Following Nahum et al. [3], we propose that $\partial S \partial t = \pm v_B \, \partial S/\partial x$ sets the entanglement profile for the spreading operator. We conclude that the entanglement within the spreading operator is "volume-law", but is less than that of a fully random operator. Only after the operator reaches an end of the system does it become maximally entangled.

May 1, 2017

Acknowledgments

Professor Huse, thank you for your uncanny ability to distinguish between aesthetic and empirical criteria/data. It is often argued that they are equally important in science, along the lines that beauty is truth realized, but it takes a resilient and ingenious mind not to (i) realize beauty as truth or (ii) dismiss truth based on an aesthetic deficit[1]. I am afraid that after a style that is so direct, ardent, and precise, there is only mush to come.

Professor Sondhi. I distinctively remember how you introduced us to the SG experiment in the first class of 305, how I didn't really know what angular momentum or a harmonic oscillator were, and how I completely failed to understand why anything "strange" was happening, for I had no classical expectation to run counter. Here I am, one and a half years later, still of no more significance to the universe than an oyster, but humbled to the bone (shell). Thank you for the inspiration, the flawless guidance, the dry humour, and for developing my taste for this peculiar privilege of sitting still and agitating the brain.

Bob, care-taker of the third floor, father of four, grandfather of seven, and many more things I can only fathom. Thank you for the nightly chit chats, for what eventually became a soothing Spiel of a turning lock, an emptied bin, a closed door, and for making me rum cake.

Sanjay, blackboard wizard, thank you for all the fooling around with chalks, always teaching me such wonderful things in the most effortless way.

Trithep, keyboard wizard, thank you for finding the semi-column.

Jadwin. You ask what Jadwin is? Let's say Jadwin was thought up by Dante. Apparently Dante tore his hair trying to decide where to put the wise men of ancient times[2]. According to the bible, they should have been tossed into hell. But the Renaissance conscience couldn't reconcile itself to the idea of enlightened men being packed in with all sorts of sinners and condemned to physical torture. So Dante thought up a special place for them in hell. You are, just as you were previously, in hell. But you have risen to its best and highest circle - the first circle.

[1] Was it Bacon who said, "There is not beauty without some strangeness in the proportions"?
[2] Originally, this was applied to Sharashkas.

Mum! I will send this to you, in case of another "What is this thing that you're doing called again?", scroll to the front page. No, I'm not doing Arts. Thank your for being so beautifully ignorant of this whole business here, for dealing with an eight-year old insomniac, for making me believe La Tete de Moine is a vegetable, for tickling me until I cry, and for never throwing away my bit-on apples.

Nathan, goatboy. Thank you for making me pea soup and ginger tea, for listening to my incessant rants about hypocritical Christians (not that there's anything wrong with being a hypocrite), for reading Pushkin to me until I fall asleep.

India, indispensable other. Thank you for the daisy chains, circles, letters, crab legs, 3£ piss bottles, train rides, tree trunks, lakes, Lolitas, heavy helpings of sleep, anonymous angels we've accumulated over the years. To many more years of ambiguously successful truffle hunting in Tuscany and to beyond.

To Bill Washburn, '67.

Contents

Notation

§1: **Time evolution of Local Operators**

- J, h, g - Ising couplings
  - Strongly chaotic, $(g = -1.05, h = 0.5, J = 1)$
  - Integrable, $(g = 1, h = 0, J = 1)$
- $\mathcal{O}$ - Operator
- A_i - Local operator, $i = 1, .., L$ is the site index.
- P^j - Pauli string, $j = 1, ..., 4^L$ identifies the basis element.

§2: **Spreading Dynamics**

- $C(t; i - j)$ - out-of-time-order commutator (OTOC)
- $W(t, k)$ - Pauli weight on site k
- $W(t, L)$ - Pauli weight of longest string L
- v_B, v_{PW} - Butterfly velocity, Pauli weight velocity

§3: **Entanglement: Conceptual Framework**

- $|\mathcal{O}\rangle$ - Braket notation for operators
- $\rho_{AB} = |\mathcal{O}_{AB}\rangle\langle\mathcal{O}_{AB}|$ - Superoperator, Density matrix on full system.
- $\rho_A = Tr_B(\rho_{AB})$ - Superoperator, reduced density matrix on A.
- $|P\rangle$ - Pauli basis
- $|S\rangle$ - Schmidt basis
- $\sigma_{S_{AB}}$ - Eigenvalues of ρ_A and ρ_B
- $\sqrt{\lambda_{S_{AB}}}$ - Schmidt values of ρ_{AB}
- $S(x, t)$ or $S(\rho_A)$ - Von Neumann entanglement entropy
- $S_P[R]$ - Page value for random operators

§4: **Entanglement Dynamics**

- opEE - Evolution of mid-cut entanglement entropy.
- $\partial S(t, x)/\partial t$ - The local rate of entanglement production.
- v_{opEE} - Speed at which $\mathcal{O}$ becomes fully entangled operator on the full system.
- opEEx - Evolution of operator entanglement entropy at different bipartite cuts.
- $\partial S(t, x)/\partial x$ - The spatial gradient of the entanglement across the bipartite cut as the cut is moved.

§5**: A Preliminary Explanation of Many-Body Localization**

- j_i, h_i, g_i - Coupling constants at site i.
- G - Tuning of randomness
  - RMT regime, $G \sim 1$
  - MBL regime, $G \sim 0$
- $r(t)$ - Ratio factor
- $g(t)$ - Spectral form factor

1 Introduction

Here we give an outline of the paper and summarize the main results. In §1.2, we write down the Ising model, define local operators, illustrate their time evolution, and introduce the notation of Pauli Strings. In §2, we study the spreading dynamics of Z_1, an operator initially localized on the first site, using two diagnostics, the out-of-time order commutator (§2.1) and he Pauli weight (§2.2). In the chaotic Ising model, we extract the butterfly velocity, discuss the broadening of the front of the moving operator and suggest an explanation for the persistent bias towards the identity in the Pauli weight (§2.3). In §3, we introduce the Braket notation for operators, the Schmidt decomposition (§3.5) and the von Neumann entanglement entropy (§3.6). In §4, we study the dynamics of operator entanglement. We compare the growth of the mid-cut entanglement entropy in §4.1 for three types of Hermitian operators (a local Z-operator to the left of the mid-cut, Z_{mid}, the product of L Z-operators, Z_{full}, a random product operator, $R_A \otimes R_B$) and one Unitary operator (the generator of time evolution, $U(t) = e^{-iHt}$). In §4.2, we study the evolution of the bipartite entanglement entropy at different cuts for Z_1 and relate the local rate of entanglement production with the spatial gradient of the entanglement across the bipartite cut. We further reconcile the dynamics of operator spreading and entanglement using the tensor network representation in §4.3. Finally, we conduct a preliminary study on the emergence of an MBL phase in the Ising model with random couplings (§5).

1.1 Summary of Results

Usually, we think about quantum dynamics in terms of the evolution of a state as $U(t)|\psi\rangle$. In this project, we adopt a different approach: we think of quantum dynamics in terms of *operators*. Specifically, we are interested in comparing two processes, the dynamics of operator *spreading* and the dynamics of operator *entanglement*.

The time evolution of an operator in the Heisenberg picture is of the form $U(t)\mathcal{O}U^\dagger(t)$. This alludes to the critical difference between state and operator dynamics: in the former, they are generated by a single unitary, in the latter, by two successive Unitaries. This means that the overlap between different wave functions evolved with the same unitary operator remains constant in time, $\langle\psi(t)|\phi(t)\rangle = \langle\psi|e^{iHt}e^{-iHt}|\phi\rangle = \langle\psi|\phi\rangle$. Hence, there is no notion of a state spreading. However, in the Heisenberg picture, the operator is flanked by two Unitaries, meaning that a time evolved operator $\mathcal{O}_\alpha(t)$ will generally fail to commute with another operator $\mathcal{O}_\beta(t)$ evolved under the same unitary, $[\mathcal{O}_\alpha(t), \mathcal{O}_\beta(t)] = U^\dagger(t)[\mathcal{O}_\alpha, \mathcal{O}_\beta]U(t)$. Thus, if we consider an operator that is local at a point x at time zero, $\mathcal{O}_x(0)$ we can define spreading by the extend to which it "fails to commute" with another operator at point $\mathcal{O}_y$ when evolved in time.

Recently, there has been much interest in using operator dynamics as a diagnostic of quantum chaos and as a pathway to understanding the process of quantum thermalization. This ties together the fields of quantum gravity, quantum information and condensed matter physics [2, 3, 4, 19, 24].

In quantum gravity, the dynamics of precursor operators are studied in the context of black hole physics and holography. Roughly speaking, a precursor is an operator, which when inserted a given instant of time, acts so as to reproduce the effect of an operator inserted at an earlier time. For an action of an operator at t_0 to be captured in the action of its precursor at later times, $t > t_0$, it must be that they are related by the Heisenberg evolution, $\mathcal{O}_P(t) = U(t, t_0)\mathcal{O}(t_0)U^\dagger(t_0, t)$. However, in a chaotic system, the action is scrambled[3] under time evolution, and the operator and its precursor cannot both be local. More intuitively, when expanding the Heisenberg operator using the Baker-Campbell-Hausdorff formula, we find a series of nested commutators, which can be taken as a proxy to the increasing complexity of the operator.

Precursors are popular in black hole Gedankenexperiments [2, 16]. In order to probe the geometry of a black hole, we consider a precursor $\mathcal{O}_P(t)$ in the interior. If the precursor does not add energy to the black hole, it commutes with the Hamiltonian and has no dynamics. However, if the precursor adds energy (it does not commute with the Hamiltonian), it generates a shock wave that spreads through the black hole interior. The speed at which this shock wave propagates forms an upper bound on the rate at which the black hole spreads quantum information, a phenomena known as "scrambling" [22].

[3] Will be defined shortly.

This can be viewed as a quantum version of the butterfly effect: If we treat $\mathcal{O}_P(t)$ as a local perturbation to the system, we are asking how this perturbation spreads to affect the measurement of another $\mathcal{O}_{P'}(t)$ at a distance $d_{PP'}$. The speed at which it grows is referred to as the "butterfly velocity" and traces out the boundary of the butterfly light cone in space time. Now, the phenomena of scrambling can be seen in systems much more primitive than black holes, for instance one-dimensional spatial lattices. This is the setting we are interested in.

In this project, we study the dynamics of local operators is the Ising system, defined on a one-dimensional chain with open boundaries. By tuning the couplings of the transverse and longitudinal fields, we can access both the integrable and the chaotic regimes. In the lattice setting, space is discretized and we denote operators by $A_i(t)$, where $i = 1, .., L$ is the site index and $A_i = (X_i/Y_i/Z_i, \mathrm{I}_i)$ are the usual 2×2 Pauli matrices, plus the identity element. Due to the limited system sizes accessible with exact diagonalization (ED), we start the operator at one endpoint of the chain and watch it spread toward the other end. We quantify the spreading dynamics in §2 using two diagnostics: (i) the out-of-time-order commutator $C(t; i, j)$ (OTOC) and (ii) the Pauli weight $W(t, k)$. The entanglement dynamics are the focus of §3, 4 and evaluated using the von Neumann entanglement entropy. In §5, we conduct a preliminary investigation on how the Ising system thermalizes, or fails to do so, under its own dynamics.

The OTOC (i) was recently proposed to quantify the process of operator spreading [23]. It is closely related to the squared commutator $\langle |[A_i(t), A_j]|^2 \rangle$. If $A_i(t)$ and A_j are far apart local operators at sites i and j respectively, then their squared commutator is zero at $t = 0$. As time evolves, $A_i(t)$ grows in size and complexity and eventually spreads to site j, at which point the squared commutator develops a finite value.

One definition of many-body chaos is that this commutator should grow to be large independently of the choice of the operators $A_i(0)$ and A_j and, once it is large, it should remain large for all times. The former can be understood in the classical context of the butterfly effect. The latter distinguishes the chaotic from the integrable system. While the size of the operator grows linearly for $t < L$ in both systems, they differ once the operator grows to the size of the full system: In the integrable system, the OTOC fluctuates at late time (with a period $\sim L$), in the chaotic system, it saturates. Henceforth and unless otherwise specified, we refer to the chaotic regime of the Ising model.

In the one-dimensional Ising chain of length L, the OTOC remains small for $t < |i-j|/v_B$. This is the time it takes $U(t)$ to convert $A_i(0)$ to a full set of $(X_i/Y_i/Z_i)$'s.The butterfly effect cone is defined by $t = |i - j|/v_B$ such that $C(t; i, j) = 1$. The OTOC increases rapidly once $t > |i - j|/v_B$ and obtains the value $C(t; i, j) = 2$. The speed v_B defines the "butterfly cone", inside of which the OTOC is large. This should be understood as the operator growing outward at speed $v_B = 1.7$.

The Pauli weight (ii) is a more "complete" version of the OTOC, introduced by Roberts et al in [2]. This measure is conceptually very simple if we expand operators in a basis of Pauli strings (e.g. $Z_1\mathrm{I}_2X_3$) and define the length of the string as the highest site index with a non-identity element (the length of $Z_1\mathrm{I}_2X_3\mathrm{I}_4\mathrm{I}_5...\mathrm{I}_L$ is three). The Pauli weight measures the fraction of all Pauli strings of length $k = 1, ..., L$. Quantitatively, the only way it differs from the OTOC is that, for the Pauli weight, we sum over all non-identity final operators, $W(t,k) = W_X(t,k) + W_Y(t,k) + W_Z(t,k)$. Computationally, the advantage is that we can calculate the weight on the k-th site simply by iteratively tracing over the spins to the right of it, $W(t,k) = 1 - W_I(t,k)$. For short times, the weight of Pauli strings of length k grows as $t^{4(k-1)}/2(k-1)!$. A similar result was first shown by Dora and Moessner in [19] for a 1d XXZ Heisenberg chain.

The two diagnostics (i) and (ii) agree fairly well: The operator spreads linearly in time with a "leading edge" that broadens sublinearly with time and, up to a precision of two digits, we find the same butterfly velocity, $v_B = 1.7$, using either of them. However, the spreading profile obtained using the Pauli weight is smoother. More subtle processes, such as the broadening of the leading edge, can only be systematically evaluated using the Pauli weight[4]. By taking the FWHM of the operator of the derivative of the Pauli weight, dW/dt, we find that the "wave front" broadens as t^β with $\beta \approx 0.6$. This suggests that the operator spreads linearly in time with a "leading edge" that broadens sublinearly with time.

The first definition of chaos we proposed was in terms of the butterfly effect. The claim was that, for a sufficiently chaotic system, a measurement at long time should be independent of the details of the initial conditions of the system. Another definition of many-body chaos is the rapid loss of local information about the local initial state of the system. This can be quantified by studying the time evolution of the entanglement of two subsystems. In a strongly chaotic system, the two subsystems will get entangled very quickly. Having identified the leading edge of the operator as a "wave front" moving along the system at the butterfly speed, it is interesting to ask how the entanglement dynamics play in. Are some regions of the spreading operator less entangled than others? Does $U(t)$ entangle different types of operators at different speeds? How does $U(t)$, the generator of entanglement, entangle itself? Do all operators get maximally entangled at long times?

In a strongly chaotic system, there is no doubt that an initially local operator will grow to span the entire system at long times. However, it is not obvious whether a spreading operator should also grow to be maximally entangled at long times. Here, the definition of "maximal entanglement" pertains to the entanglement of a fully random operator. In this context, asking whether the operator becomes "maximally entangled" is one way to ask whether it becomes "fully random". When studying the spreading

[4]We suspect this is due to the "averaging" over all final states.

dynamics, we suggested visualizing the operators in the Pauli string basis. When studying entanglement, it is helpful to think of the operators as states in the space of all operators.

Very crudely, we can just look at a matrix, element by element, and obtain a preliminary intuition of whether it is random or not. But even quantitative measures of entanglement do little more than (systematically) probing the matrix structure of operators. A random operator is expected to have little to no special structure, while physically meaningful operators have additional structure due to their special properties. Observables are Hermitian, the generator of time evolution is unitary, and the Pauli string operators have a rigid spectral constraint (the spectrum is ± 1 and highly degenerate). By asking how random or entangled these physical operators become at large times, we are quantifying how much of the initial structure withstands the Heisenberg dynamics.

The most common measure of entanglement is the von Neumann entanglement entropy. It is known to encode universal properties of quantum ground states and has led to new perspectives on the AdS-CFT correspondance. But the dynamics of entanglement have been far less studied, especially in the context of operators. A recent treatment by Nahum et al [3] characterizes entanglement growth under random unitary dynamics for *states*. Generalizing the treatment of Nahum et al. to operators, we study the evolution of the entanglement entropy both as a function of time for a bipartite at mid-chain (opEE) and as a function of the position of the bipartite cut (opEEx).

For the opEE, we consider three types of Hermitian operators: (i) a local Z-operator to the left of the mid-cut, Z_{mid}, (ii) the product of L Z-operators, Z_{full}, (iii) a random product operator, $R_A \otimes R_B$ and (iv) the generator of time evolution, $U(t) = e^{-iHt}$. We find the following operator entanglement speeds as (i)-(iv) reach a particular entanglement entropy threshold: $v_{(i)opEE} = 0.78, v_{(ii)opEE} = 1.06, v_{(iii)opEE} = 1.24, v_{(iv)opEE} = 0.57$. The derivative of the opEE, dS/dt, for the Unitary suggests that the leading edge of the operator goes faster than the trailing edge. The state entanglement speed for a random product state $\Psi_A \otimes \Psi_B$ is $v_{EE} = 1.08$. We compare the saturation values of (i)-(iv) with that of an average over random operators and find that only those operators without a local component at $t = 0$ saturate this value.

For the opEEx, we study the entanglement profile of $Z_1(t)$. The leading edge of the opEEx is moving at the butterfly speed v_B and the entanglement entropy appears to satisfy a wave equation, $S(t, x) = \lambda(v_B t \pm x)$, λ =constant. The final entropy profile at times after which the operator has become maximally entangled is the pyramid $S(x) = 2L\min\{x, L - x\}$. We relate the local rate of entanglement production, $\partial S(t, x)/\partial t$, and the spatial gradient of the entanglement across the bipartite cut, $\partial S(t, x)/\partial x$ as the cut is moved by $\partial S/\partial t = G(\partial S/\partial x)$. Following Nahum et al. [3], we assume this is a smooth even function so that the intersection $dS/dt = \pm v_B\, dS/dx$

is the only solution for the spreading operator. We conclude that the entanglement within the spreading operator is "volume-law", but is less than that of a fully random operator. Only after the operator reaches an end of the system does it become maximally entangled. An open question is whether there are contexts in which other values of dS/dt and dS/dx arise in a physically "natural" situation for operators.

Despite being a correct and simple measure of entanglement, the von Neumann entanglement entropy does not captures the full entanglement structure of the operator. This is because when computing $S = -\sum_\gamma \lambda_\gamma \log_2 \lambda_\gamma$, we are summing over all eigenvalues of the reduced density matrix. Many nearly equal, non-zero eigenvalues give a high entropy, while a single eigenvalue equal to one is indicative of a product operator with no entanglement. While the entanglement entropy was sufficiently detailed to study some aspects of the entanglement dynamics, one can look into more detail at the spectral properties of the operators of interest.

Recall that we introduced chaos in the context of quantum gravity as a perturbation to a black hole spreading at the butterfly velocity. We then gave a second definition relevant in quantum information in terms of entanglement. The third and final definition of chaos we propose relates to the process of thermalization in quantum many-body systems, which is an important challenge in condensed matter physics.

In classical many-body systems, chaos is intimately related to thermalization. The way we generate chaos in a classical system is by rendering the classical equations of motion non-linear. This means that, in phase space, two trajectories that were initially close to each other can end up at vastly different points at later times. The time scale on which these two trajectories diverge provides a measure of this classical chaotic behavior. In a quantum many-body system, the relation between thermalization and chaos is much less clear. A common way to probe the extent to which a many-body system is quantum chaotic is to probe the level statistics of its Hamiltonian.

We make a preliminary exploration into studying the level statistics of the Ising Hamiltoniain, first by probing the nearest neighbour level spacing using the ratio factor [8] and then by studying the level repulsion between energy levels much further apart than the mean spacing using the spectral form factor [1]. We are able to demonstrate clearly that for a nonrandom chaotic chain, the level statistics are governed by random matrix dynamics and that, as we add random couplings and fields, we enter an MBL phase. We conclude that while the level statistics are a useful probe, they only give us a "yes" or "no" answer in the sense that they correspond either to a GOE distribution (RMT chaotic phase) or to a Poisson distribution (MBL phase). The understanding of the transition between the localized and delocalized phase and the question of whether the spectral form factor will be a useful tool in its investigation remains a work in progress.

1.2 Time Evolution of a Local Operator

Local Operator An operator A_j acting on a spin chain is considered *local* in $\mathcal{H}_A$ (A is a subset of sites) if it is of the form

$$A_j = ...\mathrm{I}_{i-2} \otimes \mathrm{I}_{j-1} \otimes A_j \otimes \mathrm{I}_{j+1} \otimes \mathrm{I}_{j+2}... \tag{1.1}$$

where the operator A_j is a local Pauli operator on site j and I is the identity acting on all the other sites. In general, A_j is some linear combination of Pauli operators X_j, Y_j, Z_j acting on site j.

Suppose $\mathcal{O}$ is an operator normalized in a state $|\psi\rangle$ so that $\langle\psi|\mathcal{O}^\dagger\mathcal{O}|\psi\rangle = 1$ and suppose that there exists a local operator A_j such that

$$\langle\psi|\mathcal{O}^\dagger A_j \mathcal{O}|\psi\rangle \neq \langle\psi|A_j|\psi\rangle \tag{1.2}$$

If for all A_i local in $\mathcal{H}_i$,

$$\langle\psi|\mathcal{O}^\dagger A_i \mathcal{O}|\psi\rangle = \langle\psi|A_i|\psi\rangle \tag{1.3}$$

then $\mathcal{O}$ changes the state $|\psi\rangle$ locally in $\mathcal{H}_A$. Note that when we define locality in terms of expectation values, we are only considering operators that correspond to observables. We can also assess locality with the reduced density matrix ρ_A [5]. If acting with $\mathcal{O}$ does not change ρ_A, in other words, if $\rho_A(\mathcal{O}|\psi\rangle) = \rho_A(|\psi\rangle)$, then $\mathcal{O}$ changes the state locally in $\mathcal{H}_B$. Local is a special case of localized. A localized operator is local in more than one Hilbert space.

We are interested in how an initially local operator spreads and gets entangled as a function of time. In order to study these properties, we consider an Ising spin chain with transverse and longitudinal field and open boundary conditions.

Ising Hamiltonian

$$H = J\sum_{i}^{L-1} Z_i Z_{i+1} + h\sum_{i}^{L} Z_i + g\sum_{i}^{L} X_i \tag{1.4}$$

where $i = 1, ...L$ and g and h measure the strength of the transverse and the longitudinal fields respectively. If we include two-site operators as local in our definition in (1.1), then this Hamiltonian can be considered a local operator (it is the sum of local operators).

In our numerics, we will be studying $L = 6, 8, 10, 12$ and 14. We will consider two choices for the couplings: one for which the system is strongly chaotic ($g = -1.05, h = 0.5$) [7], and one for which it is integrable ($g = 1, h = 0$). We will study the dynamics

[5] We will introduce bipartitions and the reduced density matrix in §3

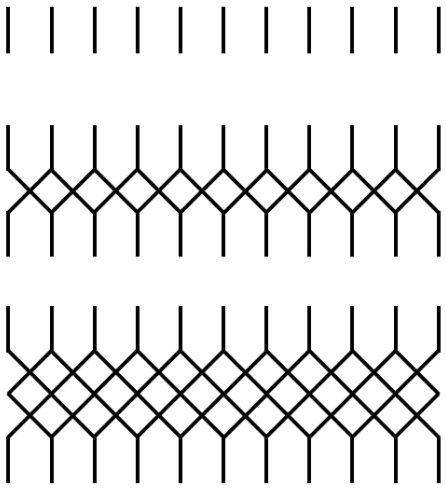


Figure 1: Tensor network representation of the time evolution operator $U(t)$ at successively larger times. At time zero, this is just the identity operator (top). At larger times, $U(t)$ acting on a local operator as in (1.5) will spread and entangle the operator (bottom panels).

of the local operators in the Heisenberg picture.

$$A_j(t) = e^{iHt} A_j(0) e^{-iHt} \tag{1.5}$$

It is helpful to visualize this pattern of growth using the tensor network geometry of the time evolution operator $U(t) = e^{-iHt}$. This is illustrated in Fig. 1 for a spin chain of eleven sites. The line endpoints correspond to the indices m_i with $i = 1, ..., 11$ and similarly for n, while the line segments are kronecker delta contractions. They are the coupling of nearest neighbouring spins and correspond to the $Z_i Z_{i+1}$ term in (1.4). The top diagram is just a representation for the identity matrix, $\mathrm{I}_{mn} = \delta_{m_1 n_1} ... \delta_{m_{11} n_{11}}$. The middle diagram corresponds to $[U(t)]_{mn}$ at very small t, when only adjacent sites are coupled. At larger times, sites further away from each other are coupled (not in a physical sense, but via entanglement). This is a graphical way of understanding how $U(t)$ spreads and entangles operators and we will come back to this representation in §4 when synthesizing the spreading and entanglement dynamics.

In this setting, when the operator begins at one end of the chain, and spreads towards the other end, the speed at which it spreads is bounded by the Lieb-Robinson bound on the commutator of local operators [5].

$$||[A_j(t), A_i]|| \leq c_0 ||A_j|| ||A_i|| e^{c_1 t - c_2 |i-j|} \tag{1.6}$$

where $||A|| = \sqrt{Tr(A^\dagger A)}$ is the infinity norm and c_0, c_1 and c_2 are constants that can (in principle) be extracted from the Hamiltonian. There are several subtleties worth noting. Firstly, the bound is valid as long as the interactions decay exponentially with distance [2]. Secondly, it produces an upper bound on the *group* velocity associated to the model under consideration. The "actual velocity" associated with a particular class of *local* observables, such as the Pauli strings under consideration, may be quite a bit less. Lastly, in the model under consideration (1.4), the estimate on the group velocity provided by the Lieb-Robinson bound will be independent of both parameters g and h. This is because they correspond to on-site terms that do not "spread". One can see this by using the interaction picture dynamics.

We will, for example, study the spreading of the operator associated to Z_1 (we have also looked at other operators to see that they behave similarly).

$$Z_1(t) = e^{iHt} Z_1(0) e^{-iHt} \tag{1.7}$$

It is helpful to think about expanding $Z_1(t)$ in a basis of *Pauli strings*, e.g. $X_1 \otimes Z_2 \otimes Y_3 \otimes X_4 ....$ which we abbreviate as $XZYZ$. Henceforth we will adopt this notation of Pauli strings by expanding (1.7) as

Pauli String

$$Z_1(t) = \sum_P^{4^L} c_P(t) P, \quad \text{where} \quad \sum_P^{4^L} |c_P(t)|^2 = 1 \tag{1.8}$$

where P denotes the 4^L dimensional Pauli basis of strings. Starting from the local operator Z_1, complicated (i.e. non-local, and, as we will discuss later on, entangled) strings such as $X_1 Z_2 Y_3 X_4 ...$ are generated as

$$e^{iHt} Z_1(0) e^{-iHt} = Z_1(0) + it[H, Z_1(0] + \frac{(it)^2}{2!}[H,[H,Z(0)]] + \frac{(it)^3}{3!}[H,[H,[H,[Z_1]]] + ... \tag{1.9}$$

where we have used the Baker-Campbell-Hausdorff expansion,

$$e^A B e^{-A} = B + [A,B] + \frac{1}{2}[A,[A,B]] + ... + \frac{1}{n!}[A,...,[A,B]]...] + ... \tag{1.10}$$

Ignoring coefficients and site indices, we find the first order term $[H, Z1(0)] = Y_1$, and the second order term $[H[H, Z_1(0)]]$ contains $X_1 Z_2, Z_1, X_1$. The sixth order term contains the Pauli strings

$$[H,[H,[H,[H,[H,[H,Z_1(0)]]]]]] = IX, XI, XX, XXX, XXXZ, XXZ, XYY, \\ XZ, XZZ, YXY, YY, YYZ, Z, ZX, ZXZ, ZZ$$

Note: we are using a notation where the operator at any site not explicitly mentioned is the local identity. Table 1 below was constructed from a numerical simulation for a system of six sites, $L = 6$, and illustrates the same growing scheme. We've chosen a strongly chaotic system with coefficients $g = -1.05$ and $h = 0.5$ and we list the coefficient, $c_P(t)$, associated with each Pauli string. We only list those coefficients with $|c_P(t)| > 10^{-6}$. At time $t = 0$, we have an operator localized on the first site. At time $t = 0.01$, we have one operator of length two, where we define the *length* of a Pauli string as the highest site index where the string contains a non-identity operator. At time $t = 0.05$, the operator has already reached the third site. Having defined a measure for the length of a Pauli string, we need to choose a quantity to measure how it spreads. This will be the focus of §2.

Unitary Time Evolution of Pauli String, L=6

c_P	P	
1	Z I I I I I	[t=0]
-0.000105	X I I I I I	[t=0.01]
-0.000209	XZ I I I I	
-0.020997	Y I I I I I	
$+0.99978$	Z I I I I I	
-0.0016761	X I I I I I	[t=0.05]
$+1.87821 \times 10^{-6}$	XXZ I I I	
$+9.39011 \times 10^{-5}$	XY I I I I	
-0.0033529	XZ I I I I	
-0.083789	Y I I I I I	
-2.81631×10^{-6}	YY I I I I	
$+8.94021 \times 10^{-5}$	YZ I I I I	
$+0.99648$	Z I I I I I	
$+1.87771 \times 10^{-6}$	ZZ I I I I	

Table 1

2 Spreading Dynamics

In this section, we study operator spreading by time evolving Z_1, an operator localized on the first site of the chain (1.7). The limited system sizes we can access via exact diagonalization (ED) prompt us to make choices that bear minimal computational cost. We look at two different diagnostics: the trace of the square of the commutator in §2.1 (expanding this gives the so-called out-of-time-order commutator), which measures the overlap of the local Z_1 operator with another local operator A_j at another site of the chain, and the Pauli weight in §2.2, which sums this measure over all operators $A_L = (X_L/Y_L/Z_L)$. We consider two different types of systems: integrable and chaotic. We show that in both the integrable and chaotic system the operator spreads linearly in time at early time. We compare the two diagnostics and conclude that for the chaotic system, the butterfly effect negates the choice of A we have to make for the commutator diagnostic.

2.1 Out-of-time-order Commutator, $C(t; i-j)$

The simplest measure of spreading is the trace of the square of the commutator, $C(t; i, j)$, which measures the extend to which $A_i(t)$ fails to commute with another A_j.

OTOC

$$C(t; i, j) = -Tr\{\rho_\beta [A_i(t), A_j]^\dagger [A_i(t), A_j]\} = -\langle [A_i(t), A_j]^\dagger [A_i(t), A_j]\rangle_\beta \tag{2.1}$$

where $\rho_\beta = e^{-iH}/Z$ is the density matrix at temperature $\beta = 1/k_B T$. The notation highlights that we are taking a thermal average at temperature $1/\beta$. For simplicity, we consider the case where $\beta = 0$, which corresponds to averaging over all states at infinite temperature, but we expect the results to hold for finite temperatures. Furthermore, we consider only hermitian operators ($A^\dagger = A$), so the commutators are purely imaginary.

$$C(t; i, j) = -Tr\{[A_i(t), A_j]^2\} \tag{2.2}$$

A criteria of chaos is that this commutator should become large at large t independent of the choice of A_i and A_j. To see this, let's unpack (2.1)

$$\begin{aligned} C(t; i, j) = -\langle & A_i(t) A_j A_i(t) A_j + A_i A_j(t) A_i A_j(t) \\ & - A_i(t) A_j A_j A_i(t) - A_j A_i(t) A_i(t) A_j \rangle \end{aligned} \tag{2.3}$$

In terms such as $A_i(t) A_j A_i(t) A_j$, there is a forward $U(t)$ evolution followed by a backward $U^\dagger(t)$ evolution. Due to this lack of temporal ordering, we will refer to the squared

commutator in as an out-of-time-order commutator (OTOC)[6]. At large times, the last two terms approach $O(1)$ values, meaning that they take the form $\langle A_i A_i\rangle\langle A_j A_j\rangle$. The way to understand this is to consider $\langle A_j A_i(t) A_i(t) A_j\rangle$ as an expectation value of $A_i(t)$ in a state where we perturbed the thermal state $|\beta\rangle$ by the operator A_j, $\langle \beta A_i|A_j(t)^2|A_i\beta\rangle$. In the large t limit, this will just reduce to the two point function $\langle A_i(t)^2\rangle$ multiplied by the norm of the perturbed thermal state $\langle \beta A_i|A_i\beta\rangle$. This is not true for the first two terms. $\langle\beta|A_i A_j(t) A_i A_j(t)|\beta\rangle$ is not an expectation value on a state, but an inner product of a state where we first apply $A_j(t)$ then A_i with a state where we first apply A_i then $A_j(t)$. In the large t limit, the states are orthogonal, and the first two terms (2.1) are small. This is why the commutator becomes large at large t (Fig. 2). Let us examine (2.1) for $A_i = Z_1$ and $A_j = Y_j$.

$$C(t;1,j)_{Z_1,Y_j} = -Tr\{[Z_1(t), Y_j]^2\} \tag{2.4}$$

It is helpful to rescale the above expression by $1/2$ and expendaing it to be in the OTOC format.

$$C(t;1,j)_{Z_1,Y_j} = 1 - Tr\{Z_1(t) Y_j Z_i(t) Y_j\} \tag{2.5}$$

We will see that $C(t;i,j)$ is of order one in a spatial region whose size grows linearly in t. This spatial region can be thought of as a ball centered on site i with a growing radius. Inside the ball the commutator is large, far outside it vanishes. An integrable model is obtained by turning off the longitudinal field (setting $h = 0$) in (1.4). This system can be solved by mapping to a system of free spinless fermions via the Jordan-Wigner transformation [6]. This mapping relates the fermion creation and annihilation operators a_j and $a_j^\dagger$ at site j to spin operators of the form $X_1X_2...X_{k-1}Y_kI_{k+1}...I_L$ and $X_1X_2...X_{k-1}Z_kI_{k+1}...I_L$. The integrable system is very special in the sense that typically diffusive quantities such as the energy density move ballistically. Figure 3 captures this diffusive behaviour: $C(t)_{Z_1,Y_j}$ grows linearly until it approaches the size of the system, after which it oscillates with a period roughly equal to the system size.

A strongly chaotic model is obtained by setting $g = -1.05$ and $h = 0.5$ [7]. Naively, one might guess that the linear growth is specific to the integrable model, and that for the chaotic system the commutator will grow slower (diffusively). However, what we find is that the rate of spreading is linear for both the integrable and the chaotic system. Only at times equal to or greater than the system size do they behave differently: the integrable model oscillates, but the chaotic system saturates. In what follows we will be considering the chaotic system exclusively.

[6]In the literature, the acronym OTOC is reserved for the out-of-time-order *correlator* and is defined as $F(t;i,j) = \langle A_i(t)^\dagger A_j^\dagger A_i(t) A_j\rangle$. If we change the sign of the squared commutator, $C(t;i,j) = \langle|[A_i(t), A_j]^2\rangle$, then the two are then related by $C(t;i,j) = 2(1 - \mathrm{Re}F(t;i,j))$. Using this definition, it is the decay of the OTOC and the growth of the squared commutator that characterizes the growth of local operators.

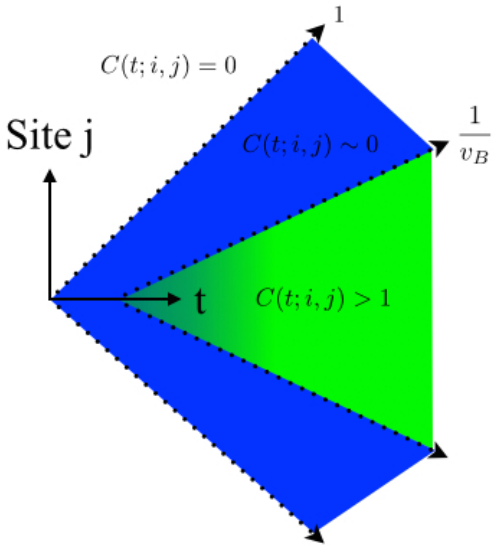
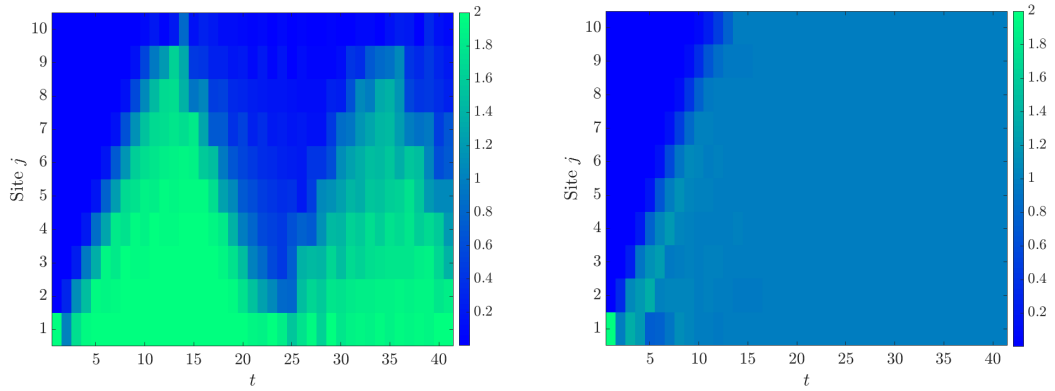


Figure 2: A schematic of the OTOC $C(t;i,j)$, which measures the extent to which two local operator $A_i(t)$ and A_j on different sites fail to commute. We illustrate the causal cone ($t = |i-j|$) and the butterfly effect cone ($t = |i-j|/v_B$). For $t < |i-j|$, the operators are space like separated and the commutator $C(t;i,j)$ must vanish by causality (white region). For $t < |i-j|/v_B$ the commutator is non-zero but negligible (blue region). Once $t \geq |i-j|/v_B$, the commutator quickly obtains the value $C(t;i,j) = 2$ (green region).



(a) Integrable system, $g = 1, h = 0$.

(b) Chaotic system, $g = -1.05, h = 0.5$.

Figure 3: Numerics of OTOC between $Z1(t)$ and Y_j for $L = 10$. For both types of evolution (integrable and chaotic), the operator spreads linearly until it approaches the size of the system. After this point, the chaotic system saturates while the integrable system oscillates.

2.2 Pauli Weight, $W(t,k)$

In this section, we generalize the measure of the OTOC in §2.1 by introducing a so-called *Pauli Weight* [2], which considers the full set of operators that have spread to site L. In other words, we decompose $Z_1(t)$ into all its constituent strings and find the weight of those with a specified length k.

Pauli Weight *The Pauli weight $W(t,k)$ is the fraction of all strings P_k of length k,*

$$W(t,k) = \sum_{P_k} |c_{P_k}(t)|^2 \tag{2.6}$$

The Pauli strings of length k have a non-identity operator at position k and only identities at positions l with $l > k$.

We begin by studying the temporal profile of the Pauli weight for the local operator $Z_1(t)$ for successively longer strings, from $k = 1$ to $k = L = 6$ in our chaotic chain ($g = -1.05, h = 0.5$). The weight of short strings initially rises, peaks, and then falls, while the weight of the longest string grows as a power law at early times and eventually saturates[7] (Fig.4). The behaviour agrees well with the commutator diagnostic in §2.1. For short times, the weight of Pauli strings of length k grows as $t^{4(k-1)}/(2(k-1))!$ (Fig. 5)[8]. This can be read off the BCH expansion: The chosen model (1.4) promotes the operator by one site only every two powers of time. If we start with Z_1 at time zero, Z appears at order t^0, Y at order t^2, XZ at t^4, XY at t^6, XXZ at order t^8 and so on. The power of $4(k-1)$ is special to this one site per two steps in the BCH expansion. Had we chosen an X_iZ_{i+1} instead of the Z_iZ_{i+1} coupling in (1.4), we would have had a power of $2(k-1)$ instead since the operator would've propagated by one site per step in the BCH expansion. This power law regime looks the same for the chaotic and the integrable system, even though they differ substantially at later times, where the chaotic system saturates and fluctuates little, while the integrable system fluctuates much more.

In Figure 6, we see how the leading edge edge of the operator passes each site as a function of time. At $t = 0$ (dark blue line in Fig. **??**), the full operator is localized on the first site. At $t = 0.5$, the operator has already spread onto the second site (see Table 1 for reference) and at $t = 1$, we get a first glance of the tail of the spreading operator. The insight gained from the spatial profile is that, in addition to spreading, the width of the moving front of the operator broadens. We will study this broadening more systematically in §2.3, where we look at the Pauli weight of operators on the full system.

[7] The sharp eye will notice that it does not saturate exactly to $W = 0.75$. This will be discussed in §2.4.

[8] Dora and Moessner showed this first for a 1D Heisenberg XXZ chain [19]. They found that at short times, the OTO commutator grows as $t^{2k}/(2k)!$.

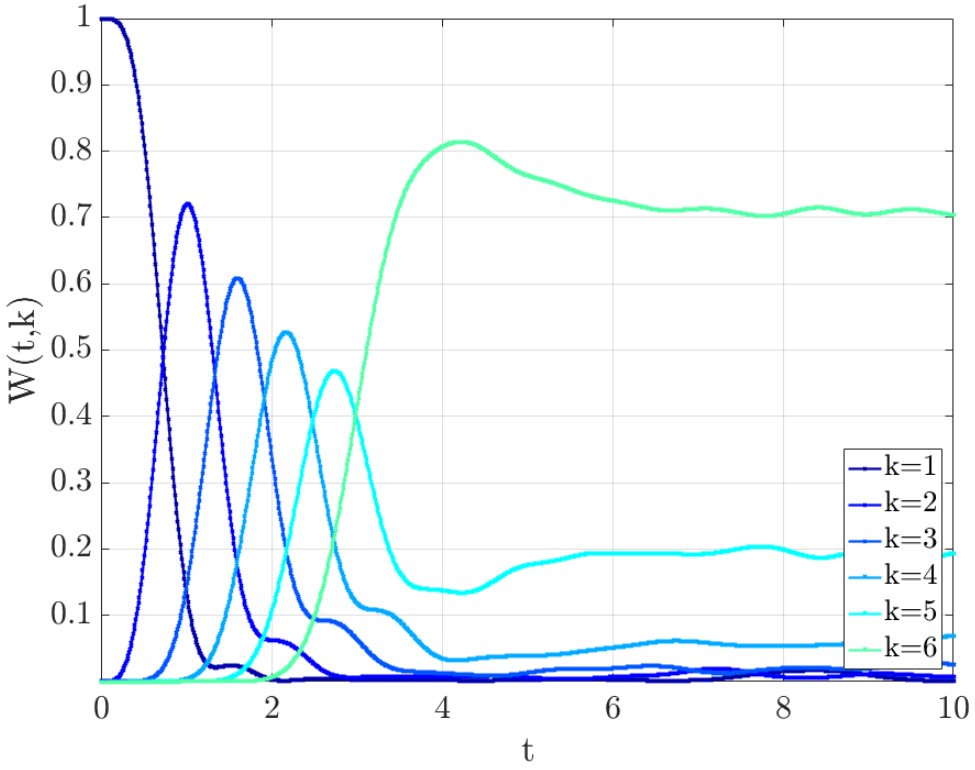
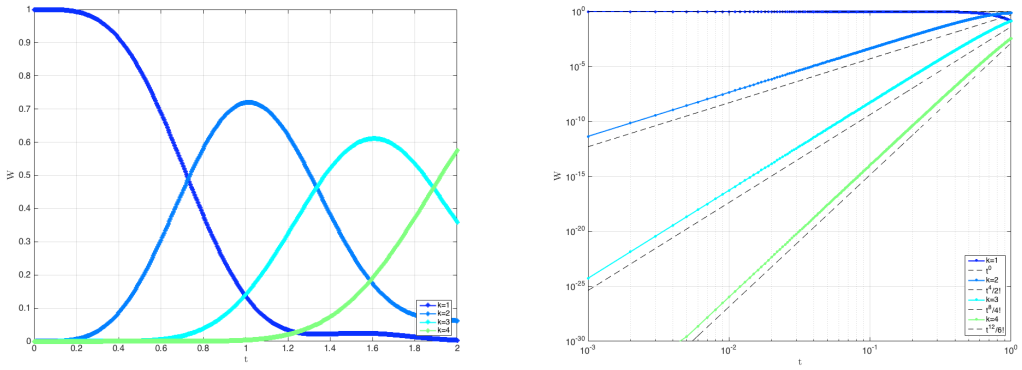


Figure 4: The Pauli weight $W(t,k)$ for $L=6$ and $g=-1.05, h=0.5$.



(a) Linear Plot

(b) Loglog Plot with Power Law Fit

Figure 5: Pauli weight $W(t,k)$ for $L=4$ and $g=-1.05, h=0.5$. The thin black dashed lines in (b) denote $t^{4(k-1)}/(2(k-1))!$.

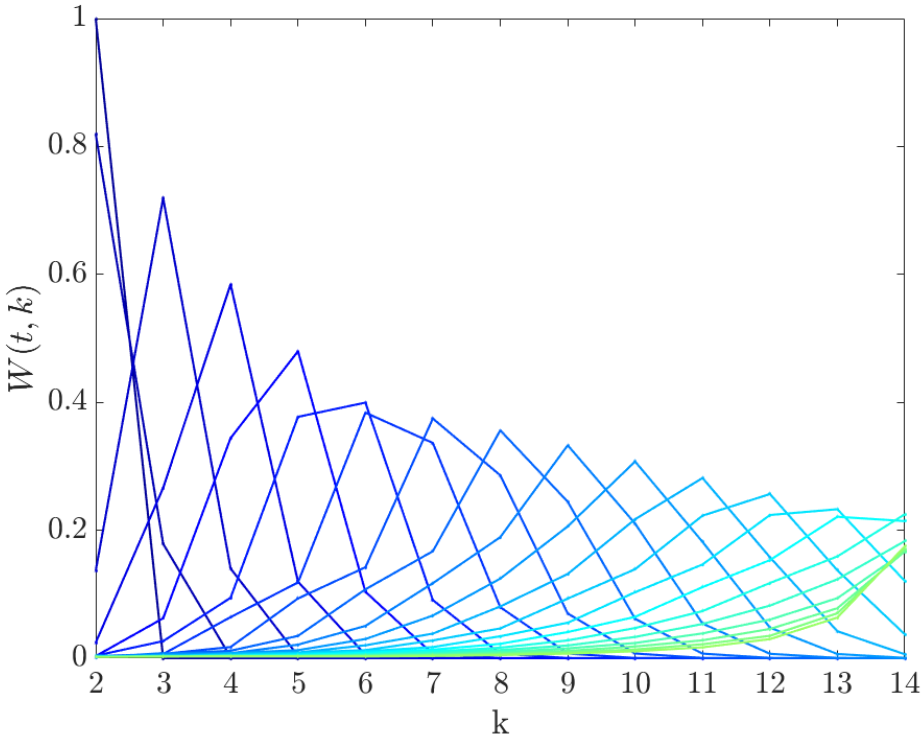


Figure 6: Spatial profile of $Z_1(t)$ for $L = 14$ on sites $k = 2$ to $k = 14$ at successive, equally spaced times from $t = 0$ to $t = 10$ in steps $dt = 0.5$. At $t = 0$ (dark blue line), the operator is localized on the first site. At late times, $t = 10$ (green line), the operator front hits a dead end and the Pauli weight saturates. In between, the leading edge of the operator is both spreading and broadening.

2.3 Pauli Weight, $W(t, L)$

We will focus mostly on the strings are those of length $k = L$ because they tell us that $Z_1(t)$ has reached the other end of the chain. Thus, unless otherwise specified, we henceforth reserve the term Pauli weight to mean the weight of all strings of length L (Fig. 7). As such, we "update" our definition of the Pauli weight as follows:

Pauli Weight L *The Pauli weight $W(t, L)$ is the fraction of all Pauli strings P_L with a non-identity operator on site L in a chain of length L.*

$$\begin{aligned} W(t, L) &= \sum_{P_L=\{X,Y,Z\}} |c_{P_L}(t)|^2 \\ &= W_{P_L=X}(t, L) + W_{P_L=Y}(t, L) + W_{P_L=Z}(t, L) \\ &= 1 - W_{P_L=\mathrm{I}}(t, L) \end{aligned} \tag{2.7}$$

The weight on the last site can be calculated conveniently by taking the trace over the last spin.

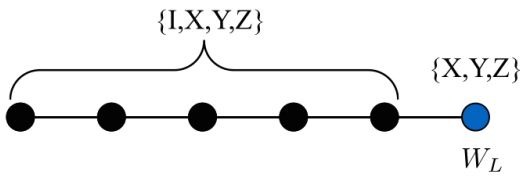


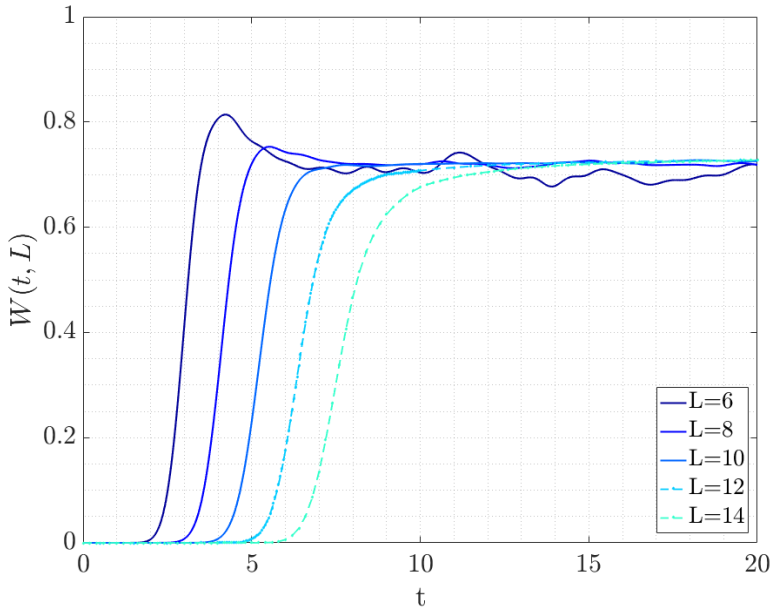
Figure 7: The Pauli weight $W(t, L)$ measures the fraction of all Pauli strings of length L. This is equal to the weight of all strings with a non-identity element on site L.

The Pauli weight illustrates how the operator spreads as a wave front, with an associated speed and width. This is shown in Figure 8(a). At early times, the operator is localized and the Pauli weight is zero. At a time proportional to the system length, the weight has a short period of rapid growth, after which it saturates. The slight overshoot for $L = 6$ and $L = 8$ is a finite size effect and goes away for the larger system sizes. The larger systems also self-average much better, as indicated by their noise-free saturation plateaus. The initial delay in the growth is due to the time it takes the leading edge of the operator to reach the other end of the chain.

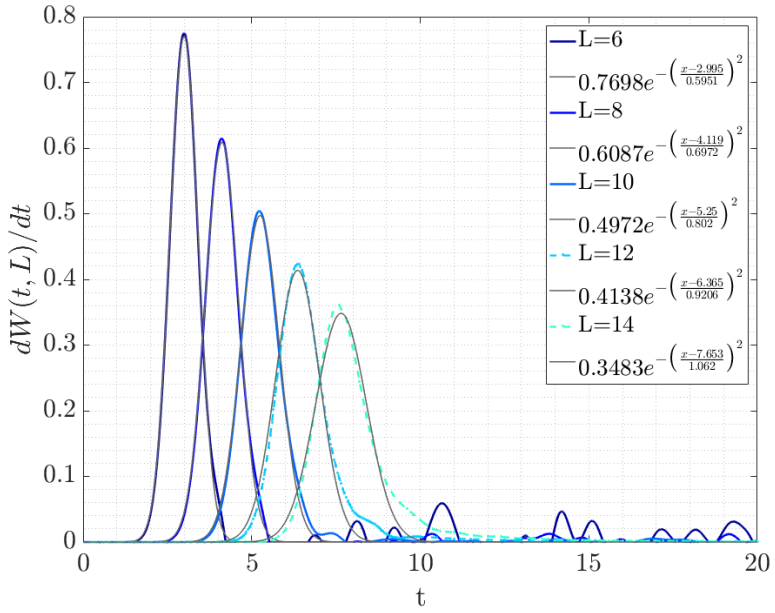
In the remainder of this section, we extract two important physical quantities from these numerics: The *speed* at which the operator spreads (the Pauli weight velocity, v_{PW}) and what appears to be a *diffusive* broadening of the leading edge of the operator. In §2.4, we address the biased saturation value of $W_{\text{saturation}} \approx 0.72$, revealing a bias in favour of the identity operator on site L that does not go away at long times and large L.

The value of the Pauli weight velocity v_{PW} depends on the choice of dynamics, but it is the same for all initially localized operators. Examples other than Z_1 that we have looked at include linear combinations of Pauli strings, such as $X_1, Y_1, X_1 + Y_1, X_1 + Y_1 + Z + 1$. We can read this velocity from Fig. 8(a) by measuring the time it takes for $W(t, L)$ to reach half its saturation value, $W_{\text{sat}}/2 \approx 0.36$, for $L = 6, 8, 10, 12, 14$. This gives us $v_{PW} = 1.690$. The broadening of the leading edge of the spreading operator is extracted by taking the FWHM of the derivative of the Pauli weight, see Fig. 8(b) and A.2. This may be consistent with diffusive $\sqrt{t}$ broadening. Note that dW/dt broadens smoothly and is close to a Gaussian shape, as expected for diffusive broadening.

We now return to the OTOC and do the analysis for the full system length L. Fig. 9 shows the wave front of the OTOC for Z_1 and Y_L. Just as we did for the Pauli weight, we extract a velocity by measuring the time it takes for $C(t, L)$ to reach half its saturation value, $W_{\text{sat}}/2 \approx 0.5$ for $L = 6, 8, 10, 12$. We find that the butterfly speed $v_B = 1.689$ derived from the OTOC agrees perfectly with the Pauli weight speed

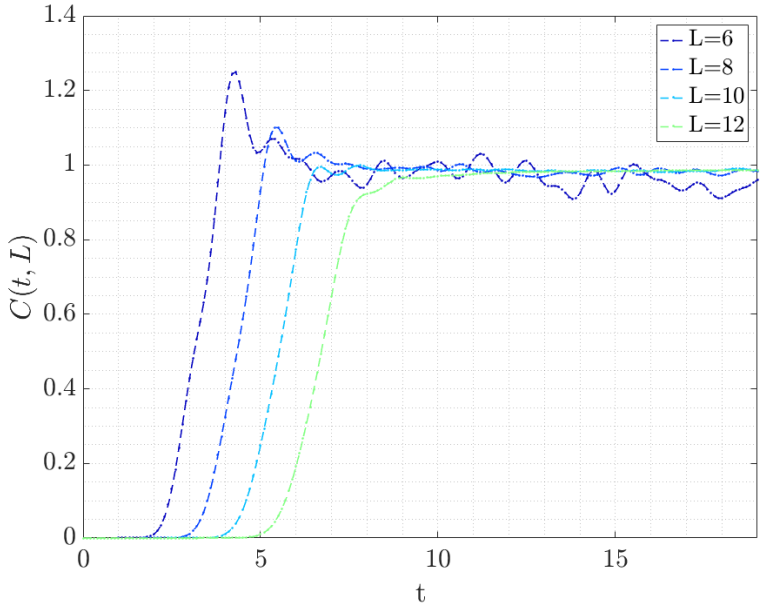


(a) Pauli Weight

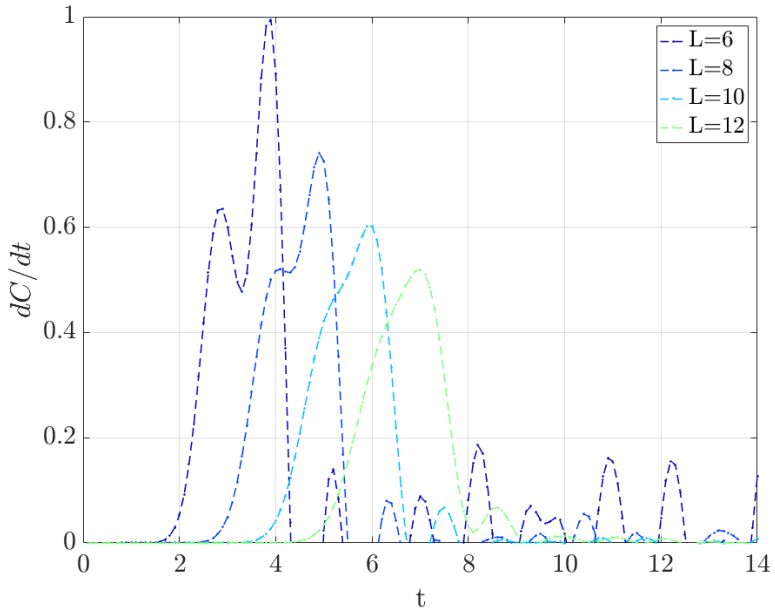


(b) Derivative of Pauli Weight with Gaussian Fit

Figure 8: Pauli Weight for initial operator $P(0) = ZI^{L-1}$ in a strongly chaotic Ising chain $g = -1.05, h = 0.5$. We consider $L = 6, 8, 10, 12, 14$ and extract $v_{PW} = 1.690$ based on the time at which the wave front at $W_{\text{saturation}}/2 = 0.36$ reaches the end of the chain in (a). The wave front broadens as $D = 0.31$ and $\beta = 0.63$ in (b). This suggests that the operator spreads linearly in time with "leading edge" that broadens sublinearly with time.



(a) OTOC



(b) Derivative of OTOC

Figure 9: OTOC for $P(0) = ZI^{L-1}$ for which we measure the overlap with Y_L in a strongly chaotic Ising chain $g = -1.05, h = 0.5$. We consider $L = 6, 8, 10, 12, 14$ and extract $v_B = 1.689$ based on the time at which the wave front at $C_{\text{saturation}}/2 = 0.493$ reaches the end of the chain in (a). The wave front broadens as and $D = 1.1$ and $\beta = 0.2$ in (b). This suggests that the operator spreads linearly in time with "leading edge" that broadens sublinearly as with time.

$v_{PW} = 1.690$. This confirms that, for a strongly chaotic system, there is no need to sum over all possible final operators. However, the OTOC front of the spreading operator broadens more weakly than the Pauli weight front. The specific structure of the leading edge is changing with L, so it is not clear if we should take the FWHM seriously. It does not look diffusive, although it might be approaching that as L is increased.

2.4 Bias towards the Identity

If $Z_1(t)$ were to become a truly random operator at large t, the Pauli weight would be expected to saturate at $W = 0.75$ since there is an equal probability of having an $(X/Y/Z/\mathrm{I})$ on the last site. What we find, however, is that the Pauli weight saturates at a slightly lower value $W_{\mathrm{sat}} \sim 0.72$, suggesting that there is a constraint involved that prevents $Z_1(t)$ from becoming fully random[9]. We can further tease out the bias by seeing which operators $(X/Y/Z)$ are underrepresented at the end of the chain due to the identity being over-represented[10]. Starting with $Z_1(0)$ or $Y_1(0)$ and taking the average over 100 random local starting operators, we find that X is at 25%, but both Y and Z are under-represented with Y being the most under-represented (the model is $ZZ + X + Z$). On the other hand, if we start with $X_1(0)$, then all three $(X/Y/Z)$ are under-represented with the maximum bias away from 25% in Y and the minimum bias in X. Furthermore, the bias towards the identity is also present for a *local* random operator of the form $R_1(t) = U^\dagger(t)R_1\mathrm{I}_2...\mathrm{I}_L U(t)$ with the same spectral constraint $(E = \pm 1)$ as the Pauli operators.

At first, we suspected that the bias might be due to the time evolution[11]. The time evolution operator $U(t) = e^{-iHt}$ at late times is a "random unitary" that is "drawn" from a different distribution than one usually means by "random unitary". If one makes a random unitary by just picking an orthonormal set of states, this random unitary does have level repulsion between its (unit magnitude complex) eigenvalues. But the spectrum of e^{-iHt} at late times wraps around the unit circle so many times that it has no level repulsion (the phases modulo 2π are Poisson distributed).

We did a simple check where we compared the Pauli weight of a random operator before and after time-evolving it. We did this over many samples to get a good average. For a 2×2 random operator localized on the first site with $E = \pm 1$, we find that the time evolution introduced a bias. However, this is not the case for a $2^L \times 2^L$ random operator (extended over the entire system) with the same spectral constraint[12]: when comparing

[9] We will find in §3.2.1 that the mid-cut operator entanglement starting with a local operator in the middle of the chain also never reaches the Page value [12] , which is also shows the operator does not get fully "scrambled".

[10] Thank you to Vedika Khemani for pointing this out.

[11] Credits to Douglas Stanford

[12] The random operator is made by choosing 2^L random orthonormal states with entries from a normal distribution and orthogonalizing them by the Graem-Schmidt procedure. We then "randomly" assign

the mean of the Pauli weight for 10^3 random operators with spectral constraint (no time evolution) with the mean of the Pauli weight of a random operator with spectral constraint over 10^3 time steps (time evolution), we find that they both saturate at $W = 0.75$. In other words, the bias is only there for the time evolution of a *local* operator with spectral constraint. This rules the time evolution as the source of the bias out.

We then realized that the culprit in making this bias was energy conservation[13]. There are certain operators that are time-independent under the time-evolution $U(t)$. These include the Hamiltonian, which is strongly biased towards identities, as well as the projection operators on to each eigenstate. For this model, the projection operators on to the ground state, the highest excited state, and on to those eigenstates nearby in energy, are local operators (since those states have low entanglement), so these operators may be well-biased towards identities. In a Floquet version of this model [11], which does not have a local Hamiltonian or low-entanglement eigenstates, there is no bias - each one of the three Pauli operators and the identity individually saturate at 25%[14]. We conclude that having the energy conservation in the Ising model produces various small-ish effects that go away when we go to the Floquet system. Furthermore, any small power of H is a conserved operator with a bias towards local identities[15].

to each state an eigenvalue of ± 1 such that the full spectrum is symmetric. $R = \sum_i^{2^L} |\phi_i\rangle E_i \langle\phi_i|$ where $E_i = \pm 1$ and $\sum_i^{2^L} E_i = Tr(R) = 0$.

[13]"We" = Prof. Huse

[14]Credits to Vedika Khemani for doing the computation.

[15]Credits to Douglas Stanford

3 Entanglement: Conceptual Framework

In this section, we lay out the formal framework necessary to understand the entanglement of operators. We first motivate the treatment of operators as vectors (wave functions) in §3.1. This will allow us to adopt the Dirac notation and write the Pauli strings in terms of kets $|P\rangle$'s and bras $\langle P|$'s. In §3.2 we discuss the two ways in which we can treat an operator on the full system, either as a matrix, or as a vector in the operator basis. We motivate the latter, paving the way to construct a basis of a composite system. In §.3.3, we find a basis for the composite system using the tensor product on the basis operators from each subspace. In §3.4, we impose that the basis in each subspace be mutually orthogonal. The existence of this basis is guaranteed by the Schmidt theorem and we will find it using the schmidt decomposition. This decomposition will be used to distinguish between seperable and entangled states in §3. and lead to the definition of the von Neumann entanglement entropy in §3.6. Finally, we give an upper bound on the entanglement entropy of states and operators in §3.7, known as the Page value [12]. A quick note on notation: we use lower indices to denote sites and upper indices to denote an element in a basis. Uppercase letters denote a subspace.

3.1 Operators in Dirac Notation

The Dirac notation for states is standard. We now motivate it for operators. Let P be an operator in a vector space V[16]. Acting on vectors it gives vectors on V, which we formally write as $P : V \to V$. We denote by $P|a\rangle$ the vector obtained by acting with P on $|a\rangle$. If $|a\rangle \in V$, then $P|a\rangle \in V$. The operator P is linear if we have $P(|a\rangle + |b\rangle) = P|a\rangle + P|b\rangle$ and $P(c|a\rangle) = cP|a\rangle$. This allows us to define

$$|Pa\rangle = P|a\rangle \tag{3.1}$$

Since a linear operator on V is also a linear operator on V^*, $P : V^* \to V^*$ and $\langle a| \to \langle a|P \in V^*$. $\langle a|P$ is the object that, acting on $|b\rangle$ gives the number $\langle a|P|b\rangle$. We can thus write operators in terms of bras and kets, written in suitable order. As an example of an operator, consider the object

$$P = |a\rangle\langle b| \tag{3.2}$$

It is a linear operator on V and on V^*. Acting on a vector, we let it act as suggested by the Dirac notation,

$$P|v\rangle \equiv |a\rangle\langle b|v\rangle \sim |a\rangle \tag{3.3}$$

[16] P should **not** be associated with a Projection operator. Instead, it should be thought of as the operator corresponding to the Pauli string P.

since $\langle b|v\rangle$ is just a number. Whether we refer to operators simply as P or as $|P\rangle$ depends on how we want to think of them, as matrices or as vectors in the operator basis. This will be addressed in the next section.

3.2 Operators as Vectors in the Operator Basis

Consider an operator $\mathcal{O}$ on the full system L. There are two ways to treat this object: As a $4^L \times 1$ vector in the operator (Pauli) basis.

$$\mathcal{O} = \sum_{P^j}^{4^L} c_{P^j} P^j = \begin{pmatrix} c_{P^1}\ P^1 \\ c_{P^2}\ P^2 \\ \vdots \\ c_{P_{4^L}}\ P_{4^L} \end{pmatrix} \tag{3.4}$$

where the P^j's are strings of length L and dimension $2^L \times 2^L$. Here, each P^j is an outer product of Paulis, one for each site.

$$P^j = P_1^j \otimes P_2^j \otimes ... \otimes P_L^j = \bigotimes_{i=1}^{L} P_i^j \tag{3.5}$$

On each site, P_i^j is a linear combination of $X_i, Y_i, Z_i, \mathrm{I}_i$,

$$P_i^j = a_i^{j0}\mathrm{I}_i + a_i^{j1}X_i + a_i^{j2}Y_i + a_i^{j3}Z_i = \sum_k a_i^{jk} A_i^{jk} \tag{3.6}$$

So we can also think of $\mathcal{O}$ as a $2^L \times 2^L$ matrix,

$$\mathcal{O} = \sum_{P^j}^{4^L} c_{P^j} P_1^j \otimes P_2^j \otimes ... P_L^j \tag{3.7}$$

To be clear on the index notation: $i = 1, ..., L$ is the site index, $k = 1, 2, 3, 4$ is the 2×2 Pauli matrix basis index and $j = 1, ..., 4^L$ is the Pauli operator basis index. Both ways of representation $\mathcal{O}$ are equally valid. There is only one caveat, which has to do with decomposing $\mathcal{O}$ into a specific basis (see §3.5). The decomposition of a vector (as in (3.4)) into a tensor product of two inner product spaces is called a Schmidt decomposition, while the factorization of a matrix M into a product of an orthogonal, a diagonal and another orthogonal matrix ($M = U\Lambda V^T$) is known as a singular value

decomposition (SVD).

$$\text{Schmidt} \longrightarrow \begin{pmatrix} \vdots \end{pmatrix}_{4^L \times 1}$$

$$\text{SVD} \longrightarrow \begin{pmatrix} \cdots & \cdots & \cdots & \cdots \\ & \ddots & \ddots & \\ \cdots & \cdots & \cdots & \cdots \end{pmatrix}_{2^{\ell_A} \times 2^{L-\ell_A}}$$

The Schmidt decomposition is essentially just an SVD in disugise, but the benefit of being applicable to an operator $\mathcal{O}$ in "vector form" will become apparent in §.3.3 where we discuss bipartitions. It turns out that reshaping a matrix to accurately reflect the writing of the Hilbert space into a product space of the two subspaces is not so straightforward, so in this context, it makes more sense to think of $\mathcal{O}$ as a vector in the Pauli basis.

In §3.1, we glanced over the fact that we need our operators to be properly normalized elements of the operator space in the same way as a state is properly normalized in the Hilbert space. Using the Pauli algebra, ($Tr(A^iA^j) = 2\delta_{ij}$, $Tr(A^iA^jA^k) = \epsilon^{ijk}2i$, we can show that this is true. Consider the spreading operator in the basis of products of Pauli strings,

$$|\mathcal{O}(t)\rangle = \sum_{P^j} c_{P^j}(t)|P^j\rangle \tag{3.8}$$

where P^j is a string of Pauli operators at different sites. The inner product of such strings satisfies

$$\langle P^i, P^j\rangle = \frac{1}{2^L} Tr(P^iP^j) = \delta_{ij} \tag{3.9}$$

3.3 Bipartition

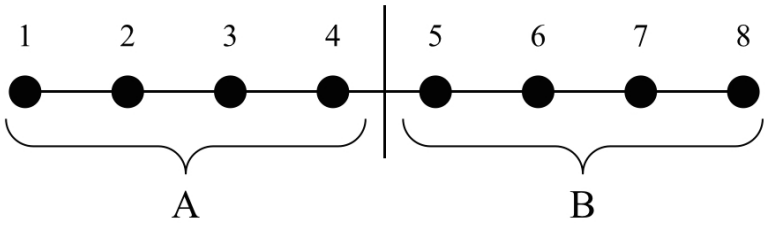


Figure 10: Bipartition

The locality of H in (1.4) means that by setting one interaction to zero, we can decompose the spin chain into two subsystems A and B, each of which is a proper quantum system by itself (Fig. 10). It is important to understand that such a bipartition is not a physical separation, instead, it means that, mentally, we are writing the Hilbert space as a product space of the two subsystems, $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B$.

One could decompose a spin chain into more than two subsystems, resulting in a multipartite system given by the tensor product $\mathcal{H}_1 \otimes \mathcal{H}_2 \otimes ... \otimes \mathcal{H}_N$ of the space corresponding to each of the subsystems. However, entanglement is simply and systematically described for bi-partitions, while for multipartions of more than two parts, a useful and systematic way of describing and analyzing the multipartite entanglement has not yet been found.

3.4 Pauli Basis

The easiest way to find a basis for a tensor product space is to use tensor products of basis operators from each sub-space. Let P_A denote the Pauli strings on subsystem A, and similarly for P_B. If A is composed of ℓ_A spins, its complement will have $\ell_B = L - \ell_A$ spins. If the operator in $\mathcal{H}_A$ and $\mathcal{H}_B$ are

$$|\mathcal{O}_A\rangle = \sum_{P_A}^{4^{\ell_A}} a_{P_A} |P_A\rangle, \qquad |\mathcal{O}_B\rangle = \sum_{P_B}^{4^{\ell_B}} b_{P_B} |P_B\rangle \tag{3.10}$$

respectively, then the resulting product operator on the full system is

$$|\mathcal{O}_{AB}\rangle = |\mathcal{O}_A\rangle \otimes |\mathcal{O}_B\rangle = \sum_{P_A}^{4^{\ell_A}} \sum_{P_B}^{4^{\ell_B}} a_{P_A} b_{P_B} |P_A, P_B\rangle \tag{3.11}$$

This holds for any choice of basis $\{|P_A\rangle\}$ and $\{|P_B\rangle\}$. The important thing is that it takes two quantum numbers to specify a basis state in $\mathcal{H}_{AB}$. Now, given an operator $|\mathcal{O}_{AB}\rangle$ on the full system, we can reduce it back to an operator acting only on subsystem A by tracing out the degrees of freedom corresponding to B. Accordingly, we define the partial trace for product operators

$$Tr_B(\mathcal{O}_{AB}) = \mathcal{O}_A Tr(\mathcal{O}_B) \tag{3.12}$$

For example, the partial trace of $|\mathcal{O}_{AB}\rangle = (Z + \mathrm{I})_A \otimes (X + \mathrm{I})_B$ over B is simply $Tr_B((Z + \mathrm{I})_A \otimes (X + \mathrm{I})_B) = (Z + \mathrm{I}) Tr(X + \mathrm{I}) = 2(Z + \mathrm{I})$.

3.4.1 Upgrading Subspace Operators

Any operator in $\mathcal{H}_A$ can be upgraded to an operator in $\mathcal{H}_{AB}$ by taking the tensor product with the identity operator in $\mathcal{H}_B$:

$$|\mathcal{O}_{AB}\rangle = |\mathcal{O}_A\rangle \otimes |\mathrm{I}_B\rangle = |\mathcal{O}_A, \mathrm{I}_B\rangle \tag{3.13}$$

$\mathcal{O}_A$ remains Hermitian when upgraded, so it is still an observable in $\mathcal{H}_{AB}$. The spectrum of $\mathcal{O}_A$ changes only insofar as that each eigenvalue has its degeneracy multiplied by a factor of $\dim(\mathcal{H}_B) = 2^{\ell_B}$. Furthermore, the operator in (3.13) acts only on subsystem A. To see this, consider $|\mathcal{O}_{AB}\rangle = Z_1 I_2 I_3$, where the subsystem A consists the first spin and the subsystem B of the second and third spin. In matrix form,

$$(Z_1)_A \otimes \big(I_2 \otimes I_3\big)_B = Z \otimes \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} Z & 0 & 0 & 0 \\ 0 & Z & 0 & 0 \\ 0 & 0 & Z & 0 \\ 0 & 0 & 0 & Z \end{pmatrix} \tag{3.14}$$

where Z is the usual 2×2 Pauli matrix. Now consider the vectors $a = (\alpha_1, \alpha_2) \in \mathcal{H}_A$ and $b = (\beta_1, \beta_2, \beta_3) \in \mathcal{H}_B$ where $\dim(\mathcal{H}_A) = 2$ and $\dim(\mathcal{H}_B) = 2^2$.

$$a_A \otimes b_B = \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} \otimes \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} = \begin{pmatrix} \alpha_1\beta_1 \\ \alpha_2\beta_1 \\ \alpha_1\beta_2 \\ \alpha_2\beta_2 \\ \alpha_1\beta_3 \\ \alpha_2\beta_3 \end{pmatrix} \tag{3.15}$$

We immediately see by acting with (3.14) on (3.15)

$$(Z_1)_A \otimes \big(I_2 \otimes I_3\big)_B (a_A \otimes b_B) = (Z_1 a)_A \otimes \big((I_2 \otimes I_3) b\big)_B \tag{3.16}$$

that the tensor product in $(Z_1) \otimes (I_2 I_3)$ acts trivially on $b \in \mathcal{H}_B$, it only mixes the components of $a \in \mathcal{H}_A$. So far, we have mostly just discussed product operators that act independently on subsystems A and B. But general operators are entangled.

3.5 Schmidt Basis

We first discuss the Schmidt decomposition in its original form as applied to states. This translates naturally into a Schmidt decomposition for operators.

3.5.1 Schmidt Decomposition for States

Given a particular pair of orthonormal bases $|\phi_\alpha^A\rangle$ in $\mathcal{H}_A$ and $|\phi_\beta^B\rangle \in \mathcal{H}_B$, any vector $|\Psi\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$ can be written in the form

$$|\Psi\rangle = M^{\alpha\beta}|\phi_\alpha^A\rangle \otimes |\phi_\beta^B\rangle = \sum_{\alpha\beta} m_{\alpha\beta}|\phi_\alpha^A, \phi_\beta^B\rangle \tag{3.17}$$

Just by looking at the indices, we see that the matrix of coefficients $M^{\alpha\beta}$ will contract the elements of $|\phi_\alpha\rangle^A$ and $|\phi_\beta^B\rangle$ in all possible combinations. The sum in (3.17) will therefore contain objects of the form $|\phi_\alpha^A\rangle \otimes |\phi_\beta^B\rangle$ where $\alpha \neq \beta$. It would be much nicer if we could replace the matrix in (3.17) with a diagonal matrix Λ^γ that only pairs those elements in $|\phi_\alpha\rangle^A$ and$|\phi_\beta^B\rangle$ with the same index. This can be done with an SVD.

$$M^{\alpha\beta} = U_A \Lambda^\gamma U_B^{-1} \tag{3.18}$$

where U_A is an $2^{\ell_A} \times 2^{\ell_A}$ real or complex unitary matrix, Λ^γ is a $2^{\ell_A} \times 2^{\ell_B}$ rectangular diagonal matrix with non-negative real numbers on the diagonal, and U_B is $2^{\ell_B} \times 2^{\ell_B}$ real or complex unitary matrix. The columns of U_A are the eigenvectors of $MM^\dagger$ and the columns of U_B are eigenvectors of $M^\dagger M$ [17]. Using (3.18), we can rewrite (3.17) as

$$|\Psi\rangle = \Lambda^\gamma |\psi_\gamma^A\rangle \otimes |\psi_\gamma^B\rangle = \sum_\gamma \sqrt{\lambda_\gamma}|\psi_\gamma^A, \psi_\gamma^B\rangle \tag{3.19}$$

Since $U_A^{-1} = U_A^\dagger$ and $U_B^{-1} = U_B^\dagger$, the columns of each of them form a set of orthonormal vectors, which can be regarded as basis vectors.

Theorem 3.1 (Schmidt Decomposition for States) *Suppose $|\Psi\rangle \in \mathcal{H}_{AB}$ is a pure state of the composite system AB. Then there exist a basis of orthonormal states $\{|\psi_\gamma^A\rangle\} \in \mathcal{H}_A$ and a basis of orthonormal states $\{|\psi_\gamma^B\rangle\} \in \mathcal{H}_B$ such that*

$$|\Psi\rangle = \sum_\gamma \sqrt{\lambda_\gamma}|\psi_\gamma^A\rangle \otimes |\psi_\gamma^B\rangle, \quad \textit{where} \quad \sum_\gamma \lambda_\gamma = 1 \tag{3.20}$$

where $\sqrt{\lambda_\gamma}$'s are non-negative real numbers. They are known as Schmidt coefficients or Schmidt values.

Here is the sketch of a proof. We can always represent $|\psi\rangle$ using the canonical basis

$$|\psi\rangle = \sum_\alpha \sum_\beta M_{\alpha\beta}|e_\alpha\rangle \otimes |f_\beta\rangle \tag{3.21}$$

[17] In the special case when M is positive semi-definite the SVD is equivalent to the eigendecomposition.

where $|e_\alpha\rangle \in \mathcal{H}_A$ and $|f_\beta\rangle \in \mathcal{H}_B$ are canonical bases in the respective subspaces. The vectors $|e_\alpha\rangle \otimes |f_\beta\rangle$ have the form

$$|e_\alpha\rangle \otimes |f_\beta\rangle = (0, ...0, 1, 0, ..., 0)^T \tag{3.22}$$

with 1 at position $\alpha\beta$ and zeros elsewhere. In this canonical basis

$$|\psi\rangle = (m_{11}, m_{12}, ..., m_{1i}, m_{21}, m_{22}, ..., m_{2i}, ..., m_{j1}, m_{j2}, ..., mij)^T \equiv \mathbf{res}(M) \tag{3.23}$$

where we have defined $\mathbf{res}(M)$ as the reshaping of the matrix M by ordering rows. Using the SVD (3.18) we get

$$|\psi\rangle = \mathbf{res}(U_A \Lambda U_B) = (U_A \otimes U_B^*)\mathbf{res}(\sigma_{\alpha\beta}\delta_{\alpha\beta}) \tag{3.24}$$

where we use the property $\mathbf{res}(ABC) = (A \otimes C^T)\ \mathbf{res}(B)$ of reshaping Now, by the canonical basis,

$$\begin{aligned} |\psi\rangle &= (U_A \otimes U_B^*) \sum_{\alpha}^{\ell_A} \sum_{\beta}^{\ell_B} \sigma_{\alpha\beta}\delta_{\alpha\beta}|\psi_{\alpha\beta}\rangle \\ &= \sum_{\gamma}^{min(\ell_A,\ell_B)} \sigma_{\gamma\gamma} U_A|e_\alpha\rangle \otimes U_B^*|f_\beta\rangle \end{aligned} \tag{3.25}$$

Since the $\sigma_{\gamma\gamma}$ are the square roots of the eigenvalues of the positive matrix $MM^\dagger$ we can write (3.1) with $\sigma_{\gamma\gamma} = \sqrt{\lambda_\gamma}$, $|\psi_\alpha^A\rangle = U|e_\alpha\rangle$ and $|\psi_\beta^B\rangle = U_B^*|f_\beta\rangle$.

3.5.2 Schmidt Decomposition for Operators

In §3.4, we said that (3.11) holds for any choice of basis. We now require that this basis be orthonormal with respect to the inner product

$$(\mathcal{O}_1, \mathcal{O}_2) = \frac{1}{4^L} Tr(\mathcal{O}_1^\dagger \mathcal{O}_2) \tag{3.26}$$

where $\mathcal{O}_1, \mathcal{O}_2 \in \mathcal{H}_{AB}$. The existence of such an orthonormal basis is guaranteed by the Schmidt decomposition theorem.

Theorem 3.2 (Schmidt Decomposition for Operators) *Suppose $|\mathcal{O}_{AB}\rangle$ is an operator of the composite system AB. Then there exist a basis of orthonormal operators $\{|S_A\rangle\}$ on A and $\{|S_B\rangle\}$ on B such that*

$$|\mathcal{O}_{AB}\rangle = \sum_{S_{AB}} \sqrt{\lambda_{S_{AB}}}|S_A, S_B\rangle, \quad \textit{where} \quad \sum_{S_{AB}} \lambda_{S_{AB}} = 1 \tag{3.27}$$

Given a composite operator $|\mathcal{O}_{AB}\rangle$, how do we find the Schmidt basis? Now that we are treating the operators as states, we can make superoperators that operate on the operators. The corresponding superoperopeator on AB is

$$\rho_{AB} = |\mathcal{O}_{AB}\rangle\langle\mathcal{O}_{AB}| = \sum_{P_A,P_B}\sum_{P_{A'},P_{B'}} c_{P_A,P_B;P_{A'},P_{B'}}|P_A,P_B\rangle\langle P_{A'}P_{B'}| \tag{3.28}$$

where $c_{P_A,P_B;P_{A'},P_{B'}} = \langle P_A,P_B|\rho_{AB}|P_{A'},P_{B'}\rangle$. Taking the partial trace over B sets $P_B = P_{B'}$ and eliminates one of the sums. This gives us the reduced density matrix on A,

$$\begin{aligned}\rho_A = Tr_B\big(\rho_{AB}\big) &= \sum_{P_B}\sum_{P_A,P_{A'}} c_{P_A,P_B;P_{A'}P_B}|P_A,P_B\rangle\langle P_{A'}P_B| \\ &= \sum_{P_A,P_{A'}}\left(\sum_{P_B} c_{P_A,P_B;P_{A'},P_B}\right)|P_A\rangle\langle P_{A'}|\end{aligned} \tag{3.29}$$

We diagonalize ρ_A and obtain

$$\rho_A = \sum_{S_{AB}} \sigma_{S_{AB}}|S_A\rangle\langle S_A| \tag{3.30}$$

This gives us the Schmidt basis $\{|S_A\rangle\}$ on A. Dagonalizing the reduced density matrix ρ_A is equivalent to doing an SVD on ρ_{AB}. This is because reduced density matrix is formed by tracing out the degrees of freedom on B, which is the same as multiplying $\mathcal{O}_{AB}\mathcal{O}_{AB}^\dagger = U_A\Lambda U_B^\dagger(U_A\Lambda U_B^\dagger)^\dagger = U_A\Lambda^2U_A^\dagger$. Instead of obtaining singular values, we have the eigenvalues of the reduced density matrix (they are related by a square root).

To find the Schmidt basis $\{|S_B\rangle\}$ on B, we repeat the procedure, taking a partial trace over A this time.

$$\rho_B = \sum_{S_{AB}} \sigma_{S_{AB}}|S_B\rangle\langle S_B| \tag{3.31}$$

So the eigenvalues of ρ_A and ρ_B are identical, namely $\sigma_{S_{AB}}$ for both. This is precisely why this decomposition is so useful. If we have the Schmidt values of an operator, we immediately know whether it is separable or not. Furthermore, if we bipartition the system unevenly into a large and a small subsystem, we can always do an eigenvalue decomposition on the reduced density matrix of the smaller system.

Theorem 3.3 *An operator $\mathcal{O}_{AB}$ can be written as a product operator, $\mathcal{O}_{AB} = \mathcal{O}_A\otimes\mathcal{O}_B$ iff its Schmidt number is equal to* 1*.*

In other words, if the Schmidt decomposition of an operator has more than one nonzero Schmidt number, the operator is entangled. The fact that a composite operator may

or may not be decomposable into a tensor product operator of the form $|\mathcal{O}_A\rangle \otimes |\mathcal{O}_B\rangle$ is unique to quantum systems. It means that there exist states in the tensor product space of physically distinct systems that are not tensor product states. These non-factorisable states are called entangled.

We mentioned in §3.3 that no good measure of entanglement has yet been found for multipartitions. This is because there is no analogue to the Schmidt decomposition for states or operators formed out of three or more subsystems.

3.6 Von Neumann Entanglement Entropy

The von Neumann entanglement entropy across a cut through the bond at position x at time t is defined as [9]

$$S(\rho_{A(x)}(t)) = -Tr\{\rho_{A(x)}(t)\log_2 \rho_{A(x)}(t)\} = -\sum_{\gamma} \lambda_\gamma(x,t)\log_2 \lambda_\gamma(x,t) \tag{3.32}$$

where the λ_γ's are the eigenvalues of ρ_A and the subsystem A contains the degrees of freedom to the left of x. The base two logarithm to give us an entropy with units of bits. Each spin in the chain can be assigned a certain number of bits. A fully entangled spin is assigned two bits.

Now, what will the von Neumann entropy of the reduced superoperator ρ_A be? If A and B are in a product operator and no entanglement is present, then $S(\rho_A) = 0$. Otherwise, the entropy will be larger the more entangled the subsystems are, being bounded from above by the maximum value by the average entanglement entropy of a random product operator. This is the Page value.

3.7 Page Value

Despite having defined the entropy in units of bits using the base two logarithm, we will revert to the natural logarithm in this scetion. The average entanglement entropy of a random *state* with a measure that is invariant under unitary transformations (i.e. Haar measure) was derived by Page to be [12]

$$\begin{aligned} S_P &= \left(\sum_{k=n+1}^{mn} \frac{1}{k}\right) - \frac{m-1}{2n} \\ &= \ln m - \frac{m}{2n} + \mathcal{O}(\frac{1}{mn}) \end{aligned} \tag{3.33}$$

where m and n are the dimensions of the Hilbert spaces $\mathcal{H}_A$ and $\mathcal{H}_B$ of the two subsystems and $m \leq n$. Consider spin $j = 1/2$ chain with L sites. The Hilbert space of each spin 1/2 has dimension two, so the dimension of the total chain is 2^L. If we bipartition the system into a smaller system A composed of ℓ_A spins, its complement will have

$\ell_B = L - \ell_A$ spins and the Page entropy will be given by

$$S_P[\Psi] = \ell_A \ln 2 - 2^{\ell_A - \ell_B - 1} + \mathcal{O}(2^{-L}) \tag{3.34}$$

For an even number of sites and a partition in the middle of the chain ($\ell_A = \ell_B = L/2$)

$$S_P[\Psi] = \frac{L}{2} \ln 2 - \frac{1}{2} + \mathcal{O}(2^{-L}) \tag{3.35}$$

The $-1/2$ deficit from the maximum possible entanglement entropy suggests that states are, on average, close but not fully maximally entangled.

An analytic calculation can be done based on an integration technique on the unitary group to show that the average entanglement entropy for a random operator R is just the Page value of a doubled system [9].

$$S_P[R] = 2\ell_A \ln 2 - 2^{2\ell_a - 2\ell_B - 1} + \mathcal{O}(4^{-L}) \tag{3.36}$$

Once more, there is a deficit from maximal entanglement. In the analysis of our numerical results for the state and operator entanglement entropies, we will compare their saturation values to the Page value.

3.8 Entanglement Spectrum

We chose to measure entanglement in terms of the von Neumann entanglement entropy, but this does not mean that there are not a number of other quantities that can capture the entanglement of a system. Another measure of entropy is given by the Renyi entropies,

$$S_\alpha(\rho_{A(x)}(t)) = \frac{1}{1-\alpha} \log_2(\rho^\alpha_{A(x)}(t)) \tag{3.37}$$

for $\alpha \geq 0$. For $\alpha < 1$, this reduces to the von Neumann entropy. In this case, $S_\infty(\rho_A) = -\log_2 ||\rho_A||$ where $||\rho_A||$ is the operator norm defined, (3.26).

We remark that the entropy of entanglement (be it the von Neumann or the Renyi entropy) is just a single number, but it should be rather obvious that more detailed information is revealed when the entire spectrum of ρ_A is considered. The collection of all Renyi entropies of ρ_A holds the same information than the spectrum of ρ_A

Recently, the full entanglement spectrum (the spectrum of ρ_A) has received a lot of attention in the context of topological systems and boundary theories, where it was shown to be a fingerprint to identify different topological order [20]. However, whether the entanglement spectrum is useful in the context of thermalization is not yet clear and is a topic of current research [21].

4 Entanglement Dynamics

In §4.1, we consider the growth of the operator entanglement entropy as a function of time, $S(t, x = L/2)$ for a bipartite at mid-chain, which we refer to as the opEE. In §4.2, we look more broadly at the growth cut as a function of the position of the bipartition cut, $S(t, x)$. This is the opEEx. We reconcile the dynamics of operator spreading and entanglement using the tensor network representation in §4.3.

4.1 Evolution of Mid-Cut Entanglement Entropy (opEE)

We study the growth of the mid-cut opEE as a function of time for system sizes $L = 6, 8, 10$ and 12. We examine three types of Hermitian operators: (i) a local Z-operator at the site immediately to the left of the mid-cut (Z_{mid}) (ii) the product of L Z-operators, one on each site (Z_{full}), (iii) a product operator, formed out of two random operators, each one acting on subsystem A and B, ($R_A \otimes R_B$) and one Unitary operator (iv) the time evolution operator $U(t) = \exp -iHt$. For comparison' sake, we also examine (v) the *state* entanglement entropy of a random product state on system A and B, ($\Psi_A \otimes \Psi_B$).

We refer exclusively to the von Neumann entanglement entropy (3.32) and measure it in units of bits. In all cases, we find an initial growth that is linear followed by an extensive saturation value, which approaches $S(t = \infty, x) = L - \alpha$ for the operators and $S(t = \infty, x) = L/2 - \beta$ for the random product state. Here α and β are the order one deficit, which are numbers larger than the "Page deficit". The distinct types of operators differ rather substantially in the speed at which they reach full saturation. This is summarized in Table 2, which gives the speed at which the different operators reach a threshold entropy (as explained below) of $S_{th} = 9$ [bits] and the state of $S_{th} = 4.5$ [bits]. What we find is that the random product operator has the highest entanglement speed, followed by the fully-spread (but not yet entangled) Z_{full}. Z_{mid} needs to spread before it can generate entanglement and therefore has the lowest entanglement speed of the three Hermitian operators. Furthermore, the speed at which the unitary generates entanglement is close to half the speed at which the random product operator gets entangled.

In the remainder of §4, we analyze the individual opEE profiles of the various operators, focusing on the linear growth regime and comparing their saturation values to that of an average over random operators with Gaussian spectrum on the full system. These random operators lack all the "special" properties of the operators (i)-(iv): they are neither local, Unitary, nor were they time evolved or have a spectral constraint. We will refer to the saturation value of the average over those random operators as the "R-value". Hence, by comparing the saturation value of the opEE for the operators (i)-(iv) with the R-value, we can make an educated guess on which one of the "spe-

cial" properties of the operators (i)-(iv) listed in Table 2 could impede the process of entanglement.

	$\mathcal{O}(0), \Psi(0)$	Properties	Spectral Constraint	S_{th}	v_{EE}
(i)	Z_{mid}	Hermitian, traceless, local	$\lambda = \pm 1$, symmetric	9	0.78
(ii)	Z_{full}	Hermitian, traceless	$\lambda = \pm 1$, symmetric	9	1.06
(iii)	$R_A \otimes R_B$	Hermitian	λ=real	9	1.24
(iv)	$U(t) = e^{-iHt}$	Unitary	$\lvert\lambda\rvert = 1$, λ complex	9	0.57
(v)	$\Psi_A \otimes \Psi_B$	State		4.5	1.08

Table 2: Entanglement speeds for Hermitian operators (Z_{mid}, Z_{full}, $R_A \otimes R_B$), the time evolution operator $U(t)$ and a product state $\Psi_A \otimes \Psi_B$ as they reach an entnaglement entropy threshold of $S_{th} = 9$ for operators and $S_{th} = 4.5$ for the state. All operators are normalized according to the trace norm, $Tr(\mathcal{O}\mathcal{O}^\dagger)/2^L = 1$. For a complete version of this Table (probing different thresholds), see B.

4.1.1 Hermitian Operators

We begin by studying the entanglement of a local $Z_i(t)$ operator. In this setting, where we are interested purely in the process of entanglement, we have to be mindful of where we localize $Z_i(t)$ at time zero. If we were to consider $Z_1(t)$, as we did in §2 for the OTOC and the Pauli weight, we would be absorbing the speed at which the operator spreads into the speed at which it gets entangled. This is because the process of entangling for $Z_1(t)$ has two steps: first the operator spreads to reach the cut at mid-chain. Then the entanglement at the cut grows. These two steps might happen at two different speeds and we would be getting an average of these speeds.

To look more purely at the rate of growth of the opEE for a spreading operator, we start Z immediately to the left of the mid-cut, which we denote by Z_{mid}. The entanglement growth for this operator is shown in Fig. 11(a). The short time linear growth is common to the system sizes $L = 6, 8, 10, 12$. At large times, Z_{mid} saturates close to $S(t) = L$. The black lines in Fig. 11(a) correspond to the R-values for different system sizes. Z_{mid} has a deficit from this R-value for all L, with the deficit being the largest for $L = 12$. Even though we do not show this in 11(a), this deficit persists even at large times (We checked this up to $t = 50$). A suggestive explanation for this deficit is the rigid spectral constraint of Z_{mid}: The spectrum is perfectly symmetric and highly degenerate. The spectrum of an operator is preserved under the Unitary time evolution, so despite becoming non-local and entangled over time, the spectrum

of Z_{mid} does not change[18]. This might be the reason why Z_{mid} does not become a fully random operator and saturating at the R-value, even at large times.

In order to extract the velocity at which the operator gets entangled (v_{opEE}), we need to measure the entropy relative to the saturation value. One way to do this is to add to the entropy profiles of the smaller system sizes $L = 6, 8, 10$ a time-independent constant of $(12 - L)$ bits. The result is shown in Fig. 11(b). In §2.2, we extracted the Pauli weight speed v_{PW} based on the time at which the wave front at $W_{saturation}/2$ reached the other end of the chain. We now employ the same method to find the entanglement speed, v_{opEE}. That is, we read off the speed based on the time at which the operator entanglement reaches a specified threshold S_{th} some $O(1)$ number of bits below the saturation. The result for different thresholds are listed in B.

To study an operator that is fully spread already (not only up to the mid-cut), we consider an initial operator that is the product of all the local Z's at each site of the chain, denoted by Z_{full}. This is somewhat of an artificial object (in the sense that it is fully spread but has zero entanglement), but it allows us to look purely at the opEE growth without mixing the process of spreading into it. Comparing it to Z_{mid}, we see that the opEE for Z_{full} grows much faster than for Z_{mid} (Fig. 11(c)). Furthermore, v_{opEE}'s extracted for Z_{full} (at different saturation thresholds) are systematically larger than those for Z_{mid} (by $\sim 30\%$), see Table 2 and appendix B. Z_{full} is unique amongst the three Hermitian operators we study in that it saturates at the R-value. This immediately rules out the hypothesis in the paragraph above that the deficit in the saturation value of the opEE for Z_{mid}, as compared to the R-value, could be a result of the $\lambda = \pm 1$ spectral constraint (Z_{mid} and Z_{full} have the same spectrum). Instead, we conjecture that it is the large number of identity elements in Z_{mid} that prevents the operator from becoming fully entangled. The identity operator does not get entangled under $U(t)$, so an operator with a large identity component at time zero might be "prevented" from getting fully entangled even at large times.

Can we construct a local operator that entangles even faster than Z_{mid}? A straightforward choice is to construct the product operator $R_A \otimes R_B$, where R_A and R_B act on either side of the bipartition. We take R_A and R_B to be normalized, random operators with Gaussian spectrum with support on A and B respectively. We then take the outer product between the two, $R_A \otimes R_B$. This is the operator at time zero. We average the entanglement at time t over such initial operators (the number of samples per system size are listed in 11(e)). Henceforth, when referring to $R_A \otimes R_B$, the average over initial operators is implied. This composite operator has the fastest entanglement growth (Fig. 11(e)) and the highest v_{opEE} (Table 2). However, as opposed to Z_{full}, it does not saturate at the R-value. Again, we argue that this is because $R_A \otimes R_B$ contains

[18]Consider two $n \times n$ matrices A and B. AB has the same spectrum as BA. Now take $A = (U^\dagger)^n$ and $B = \mathcal{O}U^n$, where we discretize time. Then $(U^\dagger)^n \mathcal{O} U^n$ and $\mathcal{O}U^n(U^\dagger)^n$ have the same spectrum.

an identity component at time zero, akin to Z_{mid}, which prevents it from getting fully entangled[19].

By comparing the opEE speeds of (i)-(iii) with the Pauli weight speeds in A.1, we find that the entanglement speed is much more robust to different thresholds. We verify this more systematically by looking at the derivative of the opEE, dS/dt, for Z_{mid}, Z_{full} and $R_A \otimes R_B$ (Fig. 13(a)-13(c)). This is much noisier than the dW/dt in §2.2, making a conclusive remark difficult. What we can say, however, is that if the entanglement front broadens, it does so to a lesser extent than the spreading (Pauli weight) front.

Finally, we note that the small "plateau" between $t = 0.5$ and $t = 1$ for Z_{mid} and between $t = 0.25$ and $t = 0.5$ for Z_{full} is a particular characteristic to having a Z_iZ_{i+1}-coupling in the Hamiltonian (1.4) and starting with a Z operator. We suspect that the temporary halt in the growth of the opEE is due to $U(t)$ being "busy" generating more $X's$ and $Y's$ (via the BCH expansion), which are needed to really kick off the rate of entanglement production.

4.1.2 Unitary Time Evolution Operator and State

In §3, we discussed in great detail how to measure the entanglement entropy of an operator. But how is entanglement generated, in the first place? Entangled operators are generated from disentangled operators by time evolving them with the unitary $U(t) = e^{-iHt}$.

The entangling ability of the unitary evolution operator $U(t)$ is depicted in Fig. 12(a). Comparing it to the opEE profile of local operators, we see that it saturates at a slower speed. Indeed, the entangling speed associated with $U(t)$ is only half that of $R_A \otimes R_B$. Another interesting property is that the leading front of the time evolution operator appears to broaden 13(d). Analogously to what we did for the Hermitian operators in the previous section, we computed the derivative dS/dt for 12(a). What we saw was a significant broadening as we went to larger L. We also found broadening for the random product state, $\Psi_A \otimes \Psi_B$ (Fig. 13(e)).

We conclude that, for the opEE of the Unitary, the leading edge of the operator goes faster than the trailing edge. This is because the leading edge is broadening, and explains the trends in B. Furthermore, we conjecture that the broadening of the leading front is a property of the time evolution operator. Since we time evolve the local operator by acting with $U(t)$ on the left and with $U(t)^\dagger$ on the right, this effect is canceled. However, for a state, we only act on the right, which induces a broadening.

[19]Removing the identity in $R_A \otimes R_B$ is equivalent to making $R_A \otimes R_B$ traceless. This is straightforward to implement, but it means that, to get a good average, we need to generate many new samples. Given the time restrictions, we will therefore simply outline the procedure. Using the identity $Tr(R_A \otimes R_B) = Tr(R_A)Tr(R_B)$, we can make R_A and R_B independently traceless. Generate and store $2^{L/2}$ random Gaussian numbers in a vector E_A for the spectrum of R_A. We want $Tr(R_A) = \sum(E_A) = 0$. Let $d = |\sum(E_A) - 0|$. We add $d/2^{L/2}$ to each element in E_A. It remains to normalize R_A such that

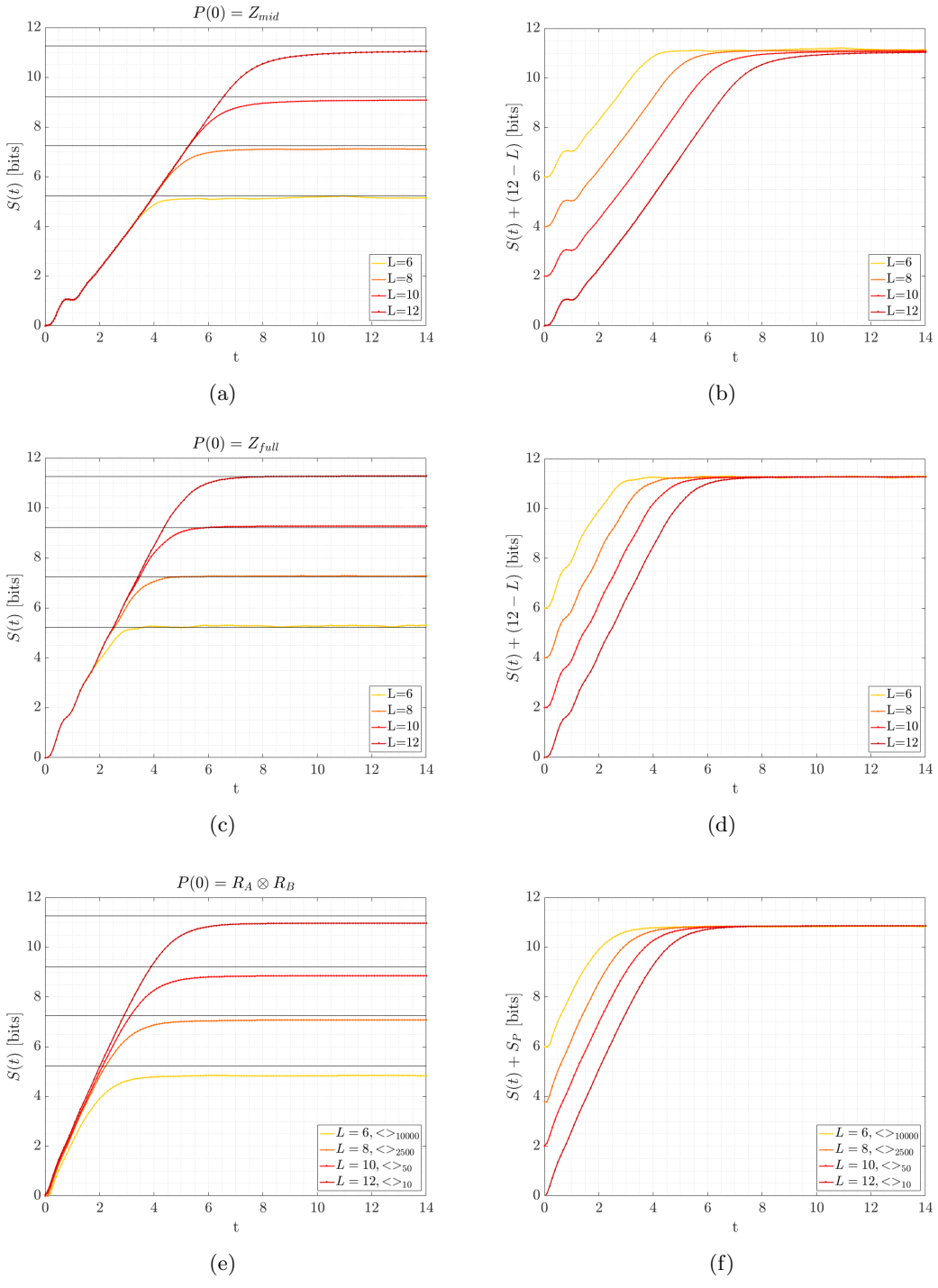


Figure 11: Growth of opEE for Hermitian operators: Z_{mid}, Z_{full} and $R_A \otimes R_B$. $(b), (d), (f)$ are measured in units of the saturation value and for (f), S_P is the difference between $S_{L=10}(\infty)$ and $S_L(t)$. The black lines correspond to the entanglement entropy of a random operator with Gaussian spectrum (no time evolution) for $L = 6, 8, 10, 12$ respectively, averaged over many such operators.

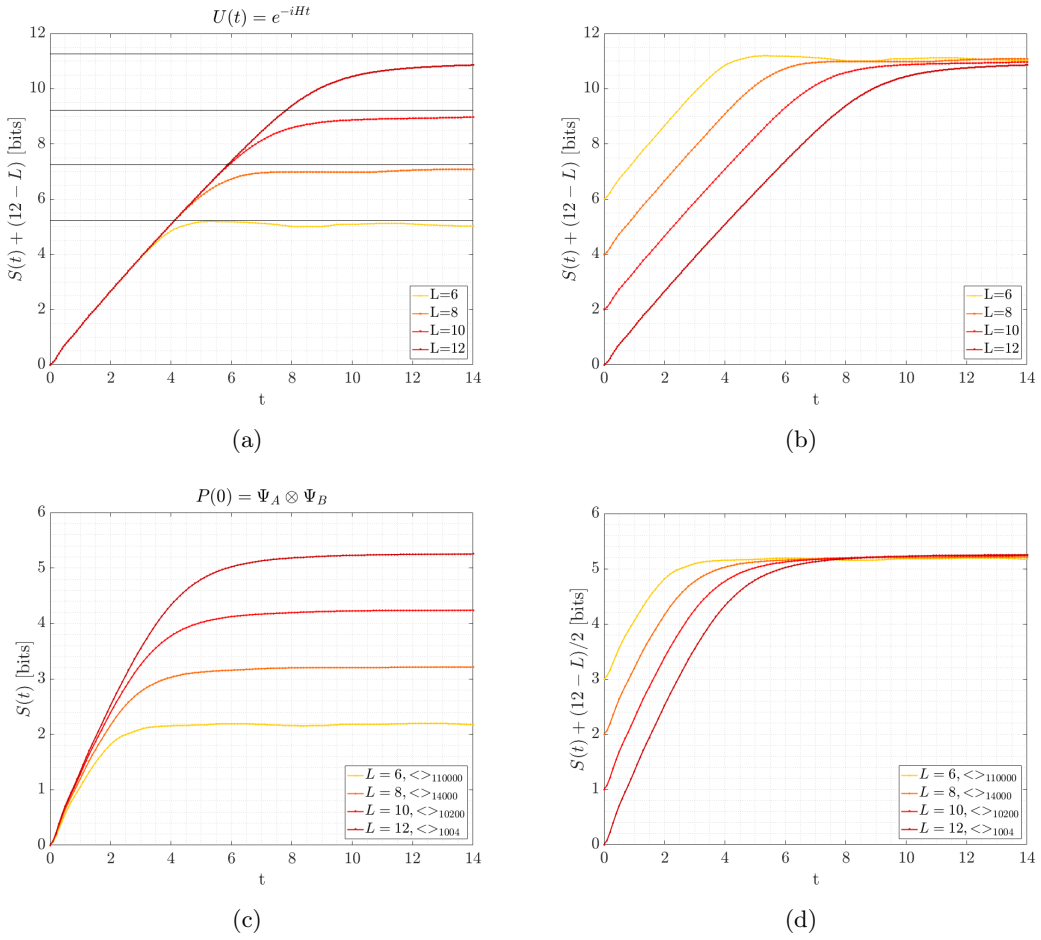


Figure 12: Growth of opEE for the time evolution operator $U(t)$ and the random product state $\Psi_A \otimes \Psi_B$ for $L = 6, 8, 10, 12$.

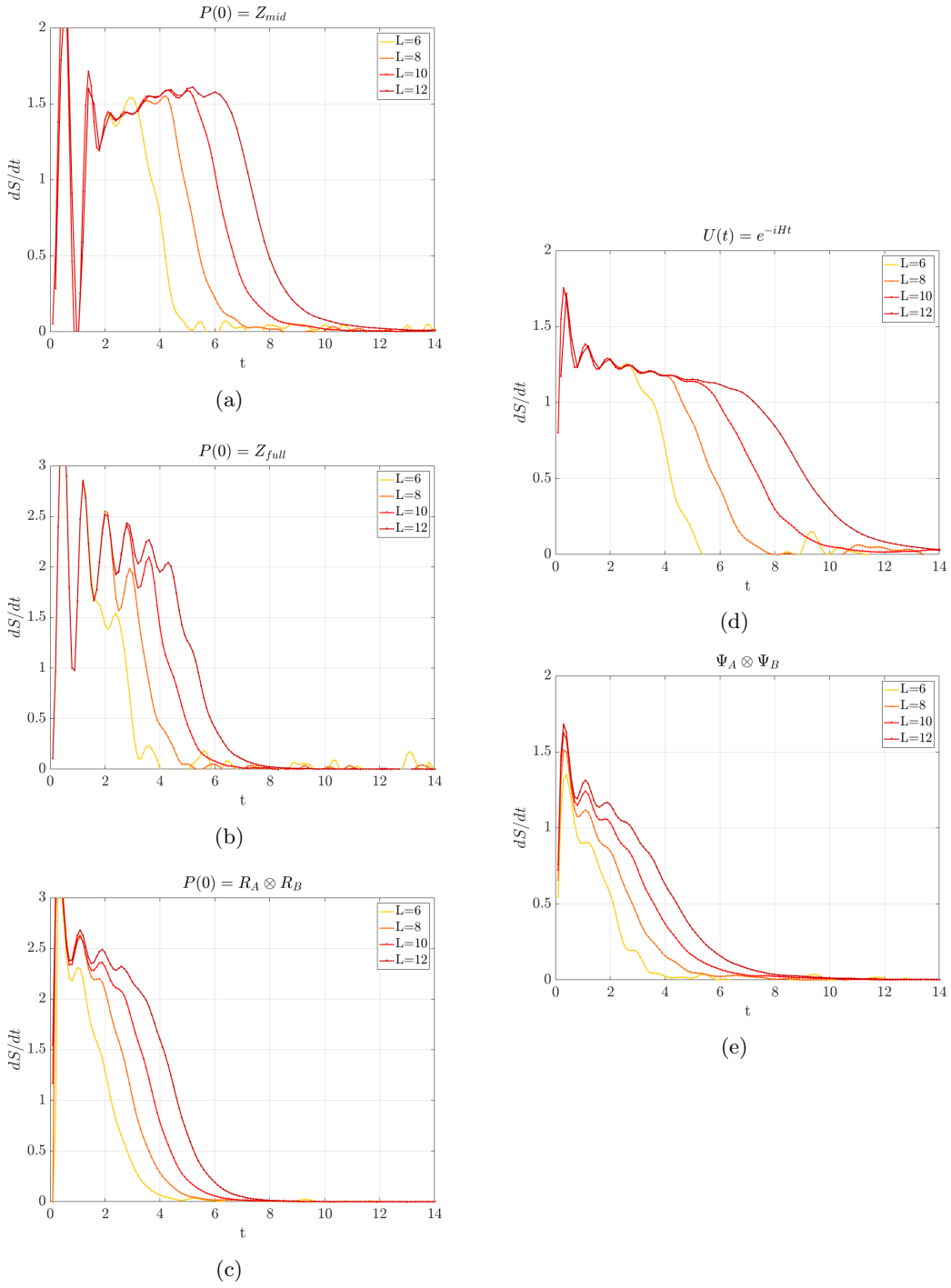


Figure 13: Derivative of the mid-cut entanglement entropy for operators (a)-(d) and a random product state (e) for $L = 6, 8, 10, 12$. In (a)-(c), there is no broadening. In (d) and (e), the leading edge of the operator goes faster than the trailing edge. We suggest that the broadening of the leading front is a property of the time evolution operator.

4.2 Evolution of Bipartite Entanglement Entropy (opEEx)

In this section we examine the evolution of the bipartite von Neumann entanglement entropy $S(t, x)$ at different cuts. A similar study was done for states in [3]. States differ from operators in that they lack an identity element that does not get entangled under the action of our e^{-iHt} if we consider only states that are at energies corresponding to nonzero temperature. An operator $\mathcal{O}$, on the other hand, can have arbitrarily many identities in its product expansion. The identity operator is "immune" to time evolution and does not get entangled.

It follows that a Pauli string with a large number of consecutive identity elements has a slower rate of entanglement than a Pauli string consisting largely of $(X/Y/Z)$. Take $Z_1 = Z_1 I_2 ... I_L$ as an example: At $t = 0$, there are only identities to the right of the first site. In this region, the time evolution operator $U(t)$ cannot generate entanglement yet. It has to spread the operator before it can entangle it. This is precisely what we see in the entanglement profile of Z_1 in Fig. 15: the leading edge of the opEEx is moving at the butterfly speed v_B. The entanglement entropy $S(t, x)$ therefore appears to satisfy a wave equation.

$$S(t, x) = f(v_B t \pm x) \tag{4.1}$$

But our numerics in §3.2 indicates that at early times $S(t, x)$ grows linearly in t at the locations where the operator is present.

$$S(t, x) = \lambda(v_B t \pm x) \tag{4.2}$$

where λ is a constant. At locations that the operator has not yet reached, $S(t, x) = 0$. At later times the operator reaches the end of the chain and then $S(t, x)$ saturates to its maximum possible value of a random operator. The final entropy profile at times after which the operator has become maximally entangled is the pyramid

$$S(x) = 2L\min\{x, L - x\} \tag{4.3}$$

Note that if we could access larger system sizes, the slopes at $t = \infty$ would be perfectly linear for x well away from $L/2$.

$Tr(R_A^\dagger R_A)/2^{L/2} = 1$.

The growth of the bipartite opEE $S(x)$ complements the mid-cut oPEE studied in §3.2. We can now relate the two physical quantities associated with entanglement: the local rate of entanglement production, $\Gamma_x \equiv \partial S(t,x)/\partial t$, and the spatial gradient of the entanglement across the bipartite cut, $\partial S(t,x)/\partial x$ as the cut is moved. Following [3], we assume that

$$\frac{\partial S}{\partial t} = G\left(\frac{\partial S}{\partial x}\right) \tag{4.4}$$

is some smooth even function like in Fig. 15(b), so the intersection in Fig. 15(b) is the only solution for the spreading operator. They are related via the butterfly velocity, $dS/dt = \pm v_B\, dS/dx$, as is shown in Fig. 15(b).

In Fig. 15(b), there are two values of dS/dt that seem "natural": that of a spreading operator and the $dS/dt = 0$ of a fully entangled operator. We also realized the case $dS/dx = 0$ by using a fully nonlocal but not entangled operator, but this is in some sense not a "natural" operator. Are there contexts in which other values of dS/dt and dS/dx arise in a physically "natural" situations? The answer to this is not clear. One can ask the same question about entanglement production in states, where different values of dS/dx arise more naturally from quantum quenches that start at some adjustable initial temperature that sets dS/dx.

4.3 Minimal Cut Interpretation

A *tensor* can be represented as a multi-dimensional array of (complex) numbers. The dimensionality of this array is called the *order* of the tensor. A scalar is a tensor of order 0, vector of order 1, a matrix would be a tensor of order 2 and so on.

A tensor network (TN) is a set of tensors where some, or all, of its indices are contracted according to some pattern. Operators can be represented as the product of tensors as:

$$|\mathcal{O}\rangle = \sum_{b_1=1}^{D_1} ... \sum_{b_{L-1}=1}^{D_{L-1}} \sum_{\substack{P_1,...P_L \\ P'_1,...,P'_L}} B^l_{b_0} W^{P_1,P'_1}_{b_0,b_1} W^{P_2,P'_2}_{b_1,b_2} ... W^{P_L,P'_L}_{b_{L-1},b_L} B^r_{b_L} |P_1...P_L\rangle\langle P'_1,...,P'_L| \tag{4.5}$$

This is shown diagrammatically in Fig. 16. $B^l_{b_0}$ denotes the left boundary vector and $B^r_{b_L}$ the right boundary vector. The W's in (4.5) correspond to the blue squares in Fig. 16. They have two *physical* indices (P_i and P'_i) and, except for the boundary terms, two *virtual* or *bond* indices (b_i and b_{i+1}). The bond indices turn out to have an important physical meaning: they represent the structure of the entanglement in the operator $\mathcal{O}$.

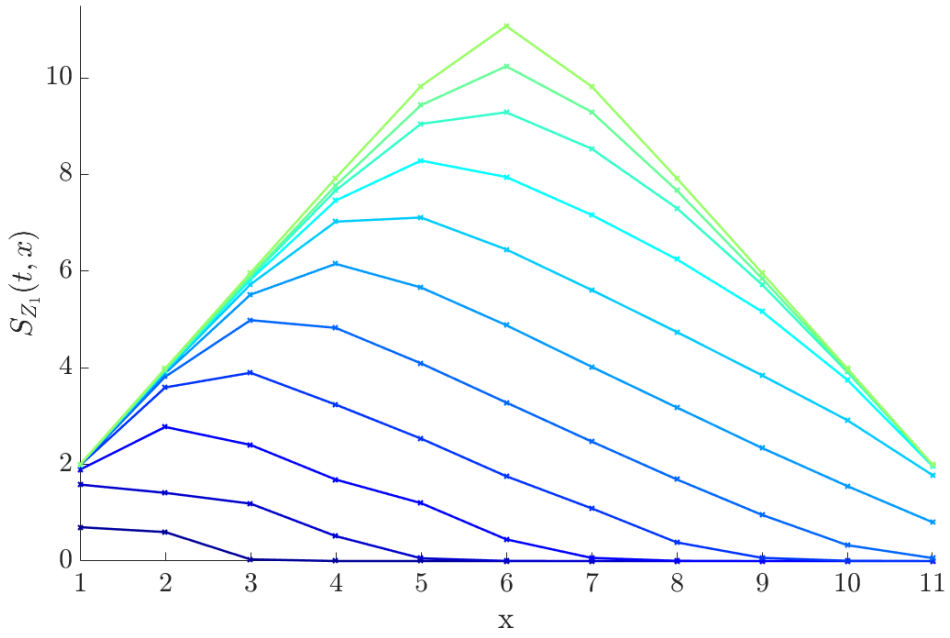
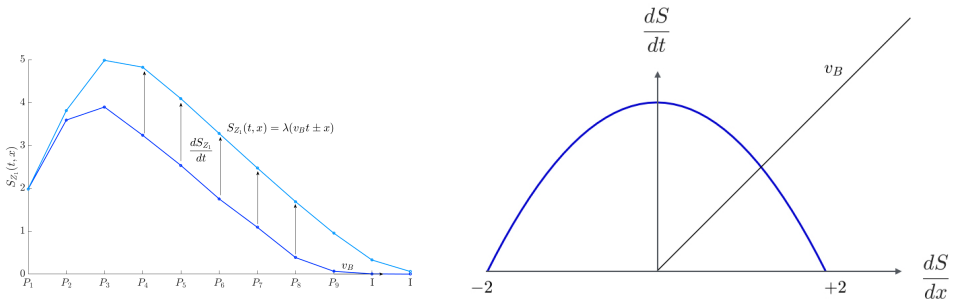


Figure 14: Growth for the entanglement $S(x)$ of Z_1 across a cut at x for $L = 12$. The lines show successive equally spaced times from $t = 1$ to 10. The top line (green) corresponds to $t = 50$.



(a) (b)

Figure 15: (a) Growth for the entanglement $S(x)$ of Z_1 across a cut at x for $L = 12$. The two lines correspond to $t = 4$ and $t = 5$. (b) The rate of entanglement production dS/dt as a 'function' of the spatial entanglement of the entanglement, dS/dx. For a initially local, then spreading operator, the two are related by the butterfly velocity v_B. The production rate dS/dt should be interpreted as follows: At $|dS/dx| = 2$, the operator is maximally entangled and the rate of entanglement production is zero. At $dS/dx = 0$ and the rate at which entanglement is produced is maximal.

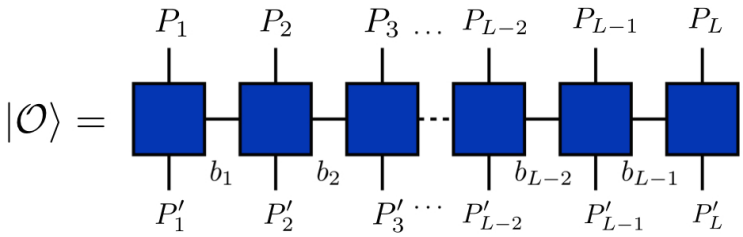


Figure 16: Matrix Product Operator (MPO) in Pauli Basis. The P_i and P_i's represent physical indices. The b_i's represent (virtual) bond indices and their number of possible values is the bond dimension D_i. For a particular cut at b_i, if $D_i = 1$, there is no entanglement present in $|\mathcal{O}\rangle$ and it can be written as a product operator. If $D_i > 1$, $|\mathcal{O}\rangle$ is entangled. If $D_i = 4^{x_i}$ with $x_i = \min(i, L-i)$ then $|\mathcal{O}\rangle$ is maximally entangled.

The number of possible values that each bond index b_i can take is known as the *bond dimension* and denoted by D. In terms of the Schmidt decomposition introduced in §3.5, D just corresponds to the Schmidt rank of each bond. Consider a bipartition of $\mathcal{O}$ in Fig. 16 at $b_{L/2}$ (L = even) into subsystems A (sites to the left of $b_{L/2}$) and B (sites to the right of $b_{L/2}$). If the bond dimension has a Schmidt rank of one, $D_{L/2} = 1$, then $S(\rho_A) = 0$. That is, the operator has zero entanglement and the TN is just a product operator. If $D_{L/2} > 1$, the TN is an entangled operator. If $D_{L/2} = 2^L$, then the operator is maximally entangled.

We are interested in the TN representation of operators in the Heisenberg picture, $\mathcal{O}(t) = U^\dagger(t)\mathcal{O}(0)U(t)$. This corresponds to a concatenation of the tensor networks of $U^\dagger(t), \mathcal{O}(0)$ and $U(t)$. For each bond we cut in the TN of $\mathcal{O}(t)$, we pick up an entanglement entropy that can be as large as

$$S_{b_i} = \log_2 D_i \tag{4.6}$$

If we take a path γ in the TN representation of $\mathcal{O}(t)$ that cuts through N bonds, the total entanglement entropy corresponding to this γ can be as large as

$$S_\gamma = \sum_{b_i=1}^{N} S_{b_i} = \sum_{b_i=1}^{N} \log_2 D_i \tag{4.7}$$

This tells us that in the concatenation of MPO's, the entanglement is not only a consequence of the bond dimensions D_i, but also of the geometry (the way in which the bond indices are connected) that represents the full network $\mathcal{O}(t)$.

The minimal cut interpretation for studying the entanglement dynamics of *states* was investigated by Nahum et al. [3]. Presumably, this was not extended to operators because, having only this bound (4.7), but not the actual values, adds to the difficulty

of finding the minimal cut through the unitary circuit representing the time evolution for a spreading operator moving along the system at the butterfly speed.

To understand this problem, we use the tensor network representation of the operator in the Heisenberg picture. First, consider the fully-spread $Z_{full}(t)$ operator on a system of eleven sites, $L = 11$ (Fig. 17). To represent the full network of $Z_{full}(t)$, we concatenate the networks for $U^\dagger(t)$, $Z_{full}(0)$ and $U(t)$. The red network represents $U(t)$ (forward time evolution), the yellow network $U^\dagger(t)$ (backward time evolution).

In the previous section, we argued that the entanglement entropy of a MPO across a bond b_i was given by the Schmidt rank of this bond, D_i. Since for $S_{Z_{full}}$ the bond dimension is the same for all bonds, (4.7) reduces to

$$S_{Z_{full,\gamma}} = N \log_2 D \tag{4.8}$$

It is clear that, in order to find the minimal cut through the TN of a fully spread operator, such as $Z_{full}(t)$ in Fig. 17, we simply need to minimize N.

The TN representation also provides insight on the entanglement dynamics of $Z_{full}(t)$. In §4.1.1, we found that Z_{full} has a higher opEE speed than $U(t)$. There is no immediately obvious reason for this, since both operators are fully spread and both have no entanglement at time zero. Using the TN representation in Fig. 17, we can make a heuristic argument of why the Unitary time evolution operator gets entangled slower than a fully spread, Hermitian operator. We discretize time.

$$Z_{full}(n) = (U^\dagger)^n Z_{full}(0) U^n \tag{4.9}$$

where $U^n = e^{-iHn}$. We found in §4.1 that the opEE of the Unitary grows linearly, we can use a similar argument to bound the entropy of $S_{Z_{full}}(n)$.

$$S_{Z_{full}}(n) \leq nS_U(n) + nS_{U^\dagger}(n) + S_{Z_{full}}(0) \tag{4.10}$$

Since $S_{Z_{full}}(0) = 0$ (at time zero, the operator is fully spread but has zero entanglement) and $S_U = S_{U^\dagger}$, this reduces to

$$S_{Z_{full}}(n) \leq 2nS_U(n) \tag{4.11}$$

We immediately see that the opEE of Z_{full} is bounded by twice that of $U(t)$.

It was simple to find a minimal cut through $Z_{full}(t)$ since the bond dimension was the same for each cut. However, this is not the case for an operator moving along the system at the butterfly speed. This structure is illustrated in Fig. 19 for $Z_{mid}(t)$ and $L = 11$. The network for Z_{mid} is the identity on all sites (white dots) except the central one at $L = 6$ (black dot). The network is simplified by the cancellation between e^{iHt} and e^{-iHt} (white region). If we had not inserted Z_{mid}, the cancellation

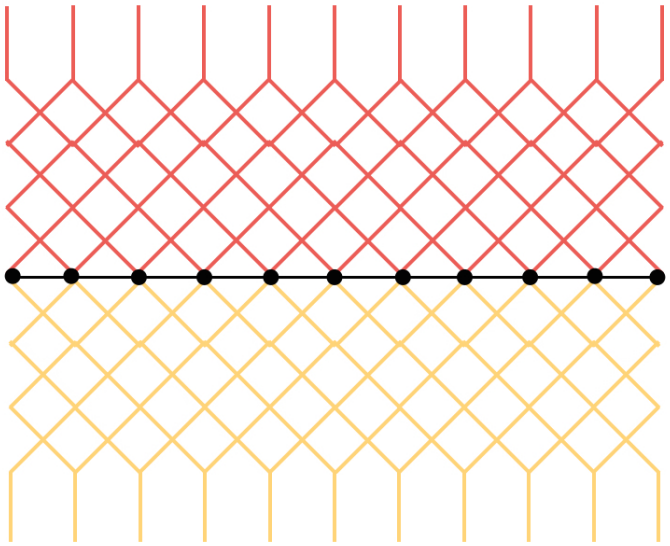


Figure 17: Tensor network representation of $Z_{full}(t)$ for $L = 11$. Each black dots corresponds to a Z Pauli. At time zero, $Z_{full}(0) = Z^L$, this operator is fully spread, but it has zero entanglement. The red network represents $U(t)$ (forward time evolution), the yellow network $U^\dagger(t)$ (backward time evolution).

would be complete and we would obtain Fig. 1. The speed at which the local operator spreads $v_B = 1.7$ (blue arrow) was found using the Pauli weight in §2.3. In the region outside of the butterfly cone, we can remove the tensor network due to the cancellation $U^\dagger(t)U(t) = \mathrm{I}$.

In §1.2, Fig. 1, we illustrated the increasing complexity of $U(t)$ pictorially using the TN representation. In §4.1, we find a quantitative measure of the speed at which $U(t)$ this process occurs. The early time regime of linear growth in the opEE of $U(t)$ (Fig. 19(a)) corresponds to the speed at which the Unitary grows to to be fully entangled, which we denote by $v_U = 0.6$. The opEEx in Fig. 19(b) reiterates that, in terms of entanglement cost, it is cheapest to cut through the circuit to either side of the mid-cut (this minimizes N in (4.7)). What is not clear is whether the entanglement is "uniformly" distributed within the network of the Unitary or whether the different contractions in $U(t)$ correspond different bond dimensions, in which case the uniform shading of the interior of the butterfly cone in Fig. 18 is misleading.

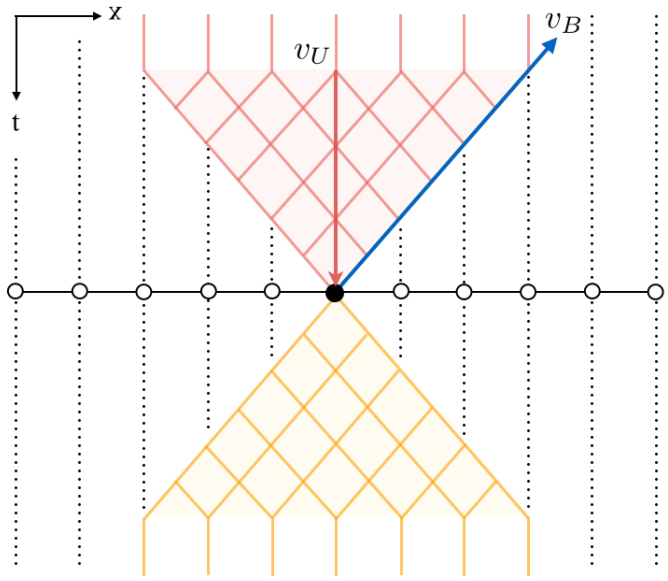
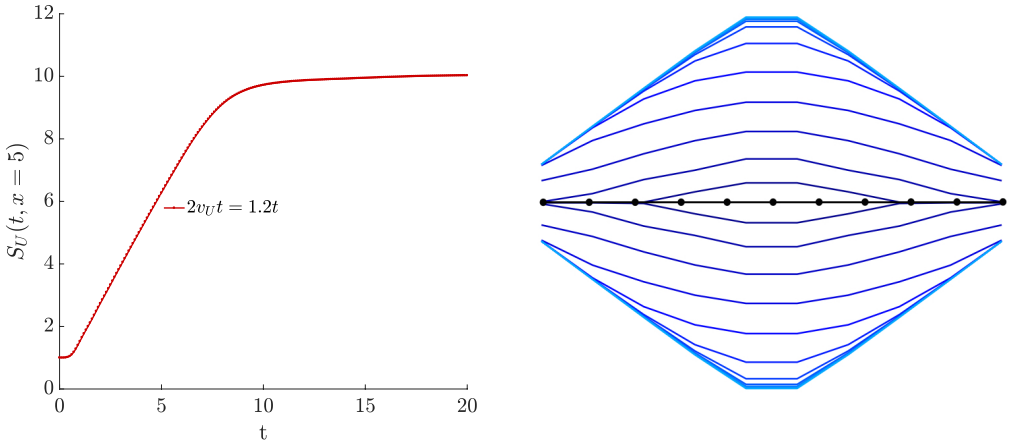


Figure 18: Tensor network representation of $Z_{mid}(t) = e^{iHt} Z_{mid}(0) e^{-iHt}$. The white circles represent identity operators and the black dot the local Z operator. The full horizontal line corresponds to $Z_{mid}(0) = \text{IIIII}Z\text{IIIII}$. The shaded area indicates the region affected by the linearly growing Z_{mid}. In the unshaded region, the forwards and backwards evolutions cancel, $U^\dagger U(t) = \text{I}$. The dotted lines indicate contractions and should be identified. $v_B = 1.7, v_U = 0.6$.



(a) Evolution of the mid-cut opEE at $x = 5$ for $S_U(t)$, $v_U = 0.6$.

(b) Evolution of the bipartite opEEx for $S_{Z_{mid}}(x)$ at different cuts, $t = 1, 2..., 9, 10, 50$.

Figure 19: Minimal Cut Interpretation for Z_{mid}, $L = 11$.

5 A Preliminary Explanation of Many-Body Localization

It is a widely held conjecture that the level statistics of nearby energy levels in quantum chaotic systems should be approximated by an appropriate random matrix ensemble [18]. This "quantum chaos" can be probed by the level statistics of the Ising Hamiltonian: A GOE[20] distribution corresponds to an Random Matrix regime (RMT), while a Poisson distribution points to a Many-Body localized phase (MBL). In order to access those two regimes by tuning the couplings, we introduce randomness to the system.

$$H = \sum_{i}^{L-1} j_i Z_i Z_{i+1} + \sum_{i}^{L} h_i Z_i + \sum_{i}^{L} g_i X_i \tag{5.1}$$

We set up the disorder conditions such that the trace of the square of the Hamiltonian is constant, i.e. $J_0^2 + H_0^2 + G_0^2 = C$ with the familiar (strongly chaotic) parameters $J_0 = 1, G_0 = -1.05$ and $H_0 = 0.5$. To keep the trace of the squared Hamiltonian constant while allowing the couplings to be random, we set

$$j_i = J_0 + a_i\big((G_0^2 - G^2)(0.5 - (G/6G_0))\big)^{1/2} \tag{5.2}$$

$$h_i = H_0 + b_i\big((G_0^2 - G^2)(0.5 - (G/6G_0))\big)^{1/2} \tag{5.3}$$

$$g_i = G + c_i\big(((G_0^2 - G^2)(G/(3G_0))\big)^{1/2} \tag{5.4}$$

with a_i, b_i, c_i independent standard Gaussian random numbers. At G near (but slightly less than) G_0, this adds equal randomness to all three terms, while as G goes to zero, the randomness is all put on the first two terms, (5.2) and (5.3), so we turn off all the transverse fields. This tuning of the couplings results in the following regimes:

- $G = |G_0| = 1.05 \sim$ couplings have no randomness, chaotic system, no level crossing, level repulsion.
- $G \sim 1 \sim$ couplings have little randomness, chaotic system, no (small) level crossing, level repulsion.
- $G \sim 0 \sim$ couplings have high randomness, ordered system, level crossing, no level repulsion.

We examine the nearest-neighbour spacing statistics using the ratio factor in §5.1 and the level repulsion between energy levels much further apart than the mean spacing using the spectral form factor in §5.2.

[20] The Ising Hamiltonian has TRS

5.1 Ratio Factor

The ratio factor is introduced in [8] as a dimensionless measure of spectral properties that takes different finite values in the thermodynamic limit $L \to \infty$, in the localized and delocalized phase and at the critical point of disorder, $W > W_c$, $W < W_c$, and $W = W_c$). Consider the level statistics between *adjacent* energy levels.

$$\delta_n = E_{n+1} - E_n \geq 0 \tag{5.5}$$

where the eigenergies of the Ising Hamiltonian (5.1) have been sorted in ascending order. The ratio of two adjacent gaps is given by [8]

$$0 \leq r_n = \frac{\min(\delta_n, \delta_{n-1})}{\max(\delta_n, \delta_{n-1})} \leq 1 \tag{5.6}$$

The numerics for $L = 10$ averaged over 200 samples are shown in Fig.20.

For $G = G_0$, $W^2 = 0$ and there is no randomness in (5.1). This means that we have an extra symmetry under reflection about the middle of the chain, $[H_{W^2=0}, R_{L/2}] = 0$. The eigenstates break up into two symmetry sectors, odd and even under $R_{L/2}$, but there is no level repulsion between the two sectors.

For large G, but $G < G_0$, the disorder is small and we have strong level repulsion, indicated by the ramp in Fig.20. This is the RMT regime. The mean value of the probability distribution of the ratio (5.6) was obtained numerically in [8, (Fig.1)]:

$$\langle r \rangle_{GOE} = 0.5295 \pm 0.0006 \tag{5.7}$$

At small G, the disorder is large and we enter the MBL regime. For Poisson statistics, the probability distribution of the ratio is $P_P(r) = 2/(1+r)^2$ and the mean value is

$$\langle r \rangle_P = 2\ln 2 - 1 \approx 0.386 \tag{5.8}$$

5.2 Spectral Form Factor

It is argued that the late time horizon fluctuations in black holes are governed by random matrix dynamics characteristic of quantum chaotic system [1]. Using the Sachdev-Ye-Kitaev (SYK) as a simple model of a black hole, the discretness of the black hole spectrum can be probed using a so-called spectral form factor [1]. The SYK model [25, 26] is considered a good candidate since it is highly chaotic: at strong coupling it saturates the chaos bound [27]. Motivated by this, we probe the Ising spectrum in (5.1).

A usual diagnostic of a discrete energy spectrum is the two-point correlation function of the Hermitian operator $\mathcal{O}(t)$,

$$\begin{aligned} G(t) &= \frac{1}{Z(\beta)} Tr\{e^{-\beta H}\mathcal{O}(t)\mathcal{O}(0)\} \\ &= \frac{1}{Z(\beta)} \sum_{m,n} e^{-\beta E_m} |\langle m|\mathcal{O}|n\rangle|^2 e^{i(E_m - E_n)t} \end{aligned} \tag{5.9}$$

Where $\mathcal{O}(t)$ is expressed in the energy basis, $|n\rangle$ are the eigenstates with eigenergy E_n, and $Z(\beta)$ is the partition function.

$$Z(\beta) = Tr\{e^{-\beta H}\} \tag{5.10}$$

There is a simpler diagnostic of a discrete energy spectrum introduced in the black hole context [17], which is obtained by pushing the partition function (5.10) into the complex plane.

$$Z(\beta, t) \equiv Tr\big(e^{-\beta H - iHt}\big) \tag{5.11}$$

By analytically continuing $\beta \to \beta + it$. As we did in §2.1 for the OTOC, we simply the calculation by setting $\beta = 0$, which corresponds to averaging over all states at infinite temperature.

$$Z(t) \equiv Tr\big(e^{-iHt}\big) \tag{5.12}$$

Now, the time average of $Z(t)$ vanishes, meaning that at large times this observable fluctuates around zero. These fluctuations can be studied by considering the squared quantity [1]

$$\left|\frac{Z(t)}{Z(0)}\right|^2 = \frac{1}{Z(0)^2} \sum_{m,n} e^{i(E_m - E_n)t} \tag{5.13}$$

At large t, the terms with energies far away from each other will average to zero and only terms with $E_m = E_n$ survive.

$$\lim_{T\to\infty} \frac{1}{T}\int_0^T dt \left|\frac{Z(t)}{Z(0)}\right|^2 = \frac{1}{Z(0)^2}\sum_E N_E^2 \tag{5.14}$$

Where N_E is the degeneracy of the energy level E. If the spectrum is highly degenerate ($N_E = 1$), then the long time average is just

$$\lim_{T\to\infty} \frac{1}{T}\int_0^T dt \left|\frac{Z(t)}{Z(0)}\right|^2 = \frac{1}{2^L} \tag{5.15}$$

We are interested in a disorder-averaged version of this quantity, so analagously to [1] we define the spectral form factor $g(t)$.

Spectral form factor

$$g(t) \equiv \frac{\langle Z(t)Z^*(t)\rangle_{c,t}}{\langle Z(0)\rangle_{c,t}^2} \tag{5.16}$$

Where the subscript $_{c,t}$ denotes a disorder average over the Gaussian random numbers $c = \{a_i, b_i, g_i\}$ in (5.2)-(5.4) as well as an average over a small time window. The numerics[21] for $L = 12$ averaged over 50 samples are shown in Fig.21.

The t parameter in $g(t)$ determines the scale of the energy differences being probed. Let $t_d \sim 1$ denote the "dip time" and $t_p \sim (\delta E)^{-1}$, where δE is the level spacing, denote the "plateau time". The early time oscillations ($t < t_d$) in Fig.21 are due to the fact that, at infinite temperature, the spectral form factor is sensitive to the hard edges at both ends of the energy spectrum.

On the ramp, ($t_d < t < t_p$) $g(t)$ is sensitive to correlations between levels that are much farther apart than the mean level spacing. For $G \sim 1$, we find that $g(t)$ is linear in t in the ramp regime, which indicates that there is level repulsion (Fig.21, right panel). This is a strong suggestion that the ramp structure in the Ising model can be attributed to RMT. As we tune the model to an MBL phase, $G \sim 0$, this level repulsion goes away.

At late times ($t > t_p$), only individual energy levels are probed and $g(t)$ saturates at $1/2^L$ (5.15) independent of G. The random couplings in (5.1) appear to rattle the energy eigenvalues of H sufficiently to make the rapidly oscillating terms in (5.13) average to zero at long times (Fig.21 upper left). By taking the time average (Fig.21, lower left) we obtain a smooth function in time.

[21] A word of caution: It is not clear which features of this we should trust since we did not get the statistical error in the mean. Some of the early time oscillations for some values of G might be real and will not go away by averaging over samples, which would be indicated by error bars that are smaller than those oscillations.

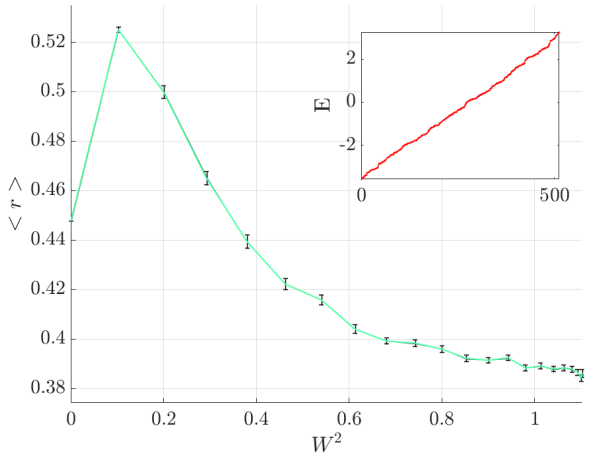


Figure 20: Ratio factor $\langle r \rangle$ for the Ising Hamiltonian (5.1) on $L = 10$ over 200 samples with the disorder strength $W^2 = (G_0^2 - G^2)$ ranging from 0 to 1. For large G, but $G < G_0$ (small disorder) we have a ramp, denoting the RMT regime. For small G (large disorder), the level statistics are Poisson distributed, denoting the MBL regime. The small panel shows the middle half of the energy spectrum.

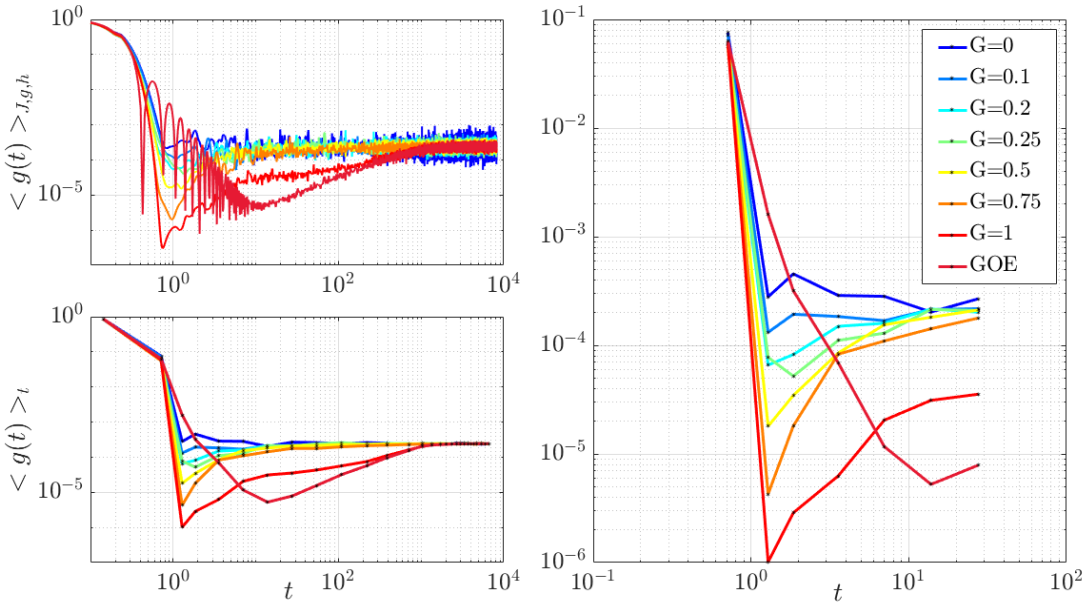


Figure 21: Spectral form factor $g(t)$ for the Ising Hamiltonian (5.1) on $L = 12$ over 50 samples with G ranging from 0 (large randomness) to 1 (no randomness), shown on loglog scale. For $G = 0$, we have the plateau, but no dip and ramp. At small G, we enter the MBL phase. At intermediate values of G, we see how the ramp develop as we turn on G. For $G = 1$, $g(t) \sim t^1$ in the ramp regime, which suggests that the ramp structure in the Ising model can be attributed to RMT. As a reference, we show $g(t)$ for a GOE.

References

[1] J.S. Cotler, G. Gur-Ari, M. Hanada, J. Polchinski, P. Saad, S.H. Shenker, D. Stanford, A. Streicher, M. Tezuka, "Black Holes and Random Matrices", arXiv preprint arXiv:1611.04650.

[2] D.A. Roberts, D. Stanford, L. Susskind, "Localized shocks", arXiv preprint arXiv:1409.8180v3.

[3] A. Nahum, J. Ruhman, S. Vijay, J. Haah, "Quantum Entanglement Growth Under Random Unitary Dynamics" arXiv preprint arXiv:1611.04650.

[4] E. Leviatan, F. Pollmann, J.H. Bardarson, E. Altman, "Quantum thermalization dynamics with Matrix-Product States" arXiv preprint arXiv:1702.08894.

[5] E. H. Lieb and D. W. Robinson, "The finite group velocity of quantum spin systems," Commun. Math. Phys. 28, 251 (1972).

[6] S. Sachdev, Quantum Phase Transitions. Cambridge Univ. Press, 2011.

[7] M. Banuls, J. Cirac and M. Hastings, Strong and Weak Thermalization of Infinite Nonintegrable Quantum Systems, Phys. Rev. Lett. 106, 050405 (2011) [arXiv:1007.3957 [quant-ph]].

[8] V. Oganesyan and D. A. Huse, "Localization of interacting fermions at high temperature", arXiv preprint arXiv:0610854.

[9] T. Zhou and D. J. Luitz, "Operator entanglement entropy of the time evolution operator in chaotic systems", arXiv preprint: 1612.07327.

[10] T. Prosen and I. Pizorn, "Operator space entanglement entropy in a transverse Ising chain," Phys. Rev. A 76, 032316 (2007).

[11] H. Kim, T. N. Ikeda, and D. A. Huse,"Testing whether all eigenstates obey the Eigenstate Thermalization Hypothesis", arXiv preprint: 1408.0535.

[12] Don N. Page, "Average entropy of a subsystem," Physical Review Letters 71, 12911294 (1993).

[13] I. Bengtsson and K. Zyczkowski, "Geometry of Quantum States" (Cambridge University Press, Cambridge, UK, 2006).

[14] O. Guhne and G. Toth, Phys. Rep. 474, 1 (2009), arXiv:0811.2803.

[15] R. Islam, R. Ma, P. M. Preiss, M. Eric Tai, A. Lukin, M. Rispoli, and M. Greiner, "Measuring entanglement entropy in a quantum many-body system," Nature 528, 7783 (2015).

[16] V. E. Hubeny, "Precursors see inside black holes", Int. J. Mod. Phys. D12 (2003) 16931698, [hep-th/0208047].

[17] K. Papadodimas and S. Raju, "Local Operators in the Eternal Black Hole," Phys. Rev. Lett. 115 no. 21, (2015) 211601, arXiv:1502.06692 [hep-th].

[18] M. L. Mehta, Random matrices, vol. 142. Academic press, 2004.

[19] B. Dora, R. Moessner, "Out-of-time-ordered density correlators in Luttinger liquids", arXiv preprint: 1612.00614.

[20] H. Li and F. D. M. Haldane, "Entanglement Spectrum of Quantum Hall States", Phys. Rev. Lett. 101 (2008), 010504

[21] S. D. Geraedts, R. Nandkishore, N. Regnault, "Many-body localization and thermalization: Insights from the entanglement spectrum", Phys. Rev. B. 93 174202 (2016).

[22] F. M. Haehl, R. Loganayagam, P. Narayan, M. Rangamani, "Classification of out-of-time-order correlators", arXiv preprint: 1701.02820.

[23] A. I. Larkin and Y. N. Ovchinnikov, "Quasiclassical method in the theory of superconductivity", Sov. Phys. JETP 28, 1200 (1969).

[24] A. Sivaramakrishnan, "Localized Excitations from Localized Unitary Operators", arXiv preprint: 1604.00965.

[25] S. Sachdev and J.-w. Ye, "Gapless spin fluid ground state in a random, quantum Heisenberg magnet," Phys. Rev. Lett. 70 (1993) 3339.

[26] A. Kitaev, "A simple model of quantum holography." Talks at KITP, April 7, 2015 and May 27, 2015.

[27] J. Maldacena, S. H. Shenker, and D. Stanford, "A bound on chaos," JHEP 08 (2016) 106, arXiv:1503.01409.

A Tables for Spreading Dynamics

A.1 Pauli Weight Speed

$P(0)$	$\triangle L$	W_{th}	v_{PW}
$Z_1 = Z \otimes I^{L-1}$	$6, 8, 10, 12, 14$	0.2	1.78
		0.3	1.72
		0.36	1.70
		0.4	1.66
		0.5	1.58
	$7, 9, 11, 13$	0.2	1.78
		0.3	1.73
		0.36	1.71
		0.4	1.70
		0.3	1.62

A.2 Broadening

dW_{Z_1}/dt		
	L	From Raw Data: d =FWHM
Dt^β, $D = 0.3081$, $\beta = 0.6301$	6	0.9741
	8	1.1346
	10	1.2948
	12	1.4606
	14	1.6476
	L	From Fit: $\sqrt{\int dt \, \frac{dW}{dt}(t - t_{\mathrm{p}})^2}$, $\frac{d^2W}{dt^2}\big\vert_{t_p} = 0$
Dt^β, $D = 0.1034$, $\beta = 0.6815$	6	0.3750
	8	0.4219
	10	0.4754
	12	0.5324
	14	0.6579

B Table for Entanglement Dynamics

$P(0)$	$\triangle L$	S_{th}* [bits]	v_{opEE}
$Z_{\text{mid}} = I^{L/2-1}ZI^{L/2}$	$6, 8, 10, 12$	6.2	0.796
		6.4	0.690
		6.6	0.683
		6.8	0.678
		7	0.678
		7.2	0.748
		7.4	0.748
		7.6	0.748
		7.8	0.750
		8	0.757
		8.2	0.763
		8.4	0.765
		8.6	0.767
		8.8	0.771
		9	0.775
		9.2	0.775
		9.4	0.774
		9.6	0.767
		9.8	0.759
		10	0.747
$Z_{\text{full}} = Z_1...Z_L$	$6, 8, 10, 12$	6.2	1.11
		6.4	1.11
		6.6	1.10
		6.8	1.08
		7	1.06
		7.2	1.05
		7.4	1.05
		7.6	1.07
		7.8	1.10
		8	1.11
		8.2	1.09
		8.4	1.07
		8.6	1.07
		8.8	1.06
		9	1.06

		9.2	1.07
		9.4	1.08
		9.6	1.08
		9.8	1.08
		10	1.09
$R_A \otimes R_B$	$6, 8, 10, 12$	6.2	1.34
		6.2	1.34
		6.4	1.33
		6.6	1.31
		6.8	1.30
		7	1.29
		7.2	1.29
		7.4	1.28
		7.6	1.28
		7.8	1.28
		8	1.27
		8.2	1.27
		8.4	1.26
		8.6	1.25
		8.8	1.25
		9	1.24
		9.2	1.24
		9.4	1.24
		9.6	1.23
		9.8	1.23
		10	1.22
$U(t) = e^{-iHt}$	$6, 8, 10, 12$	6.2	0.625
		6.4	0.619
		6.6	0.612
		6.8	0.608
		7	0.606
		7.2	0.603
		7.4	0.600
		7.6	0.596
		7.8	0.592
		8	0.589
		8.2	0.587
		8.4	0.583
		8.6	0.578

		8.8	0.574
		9	0.568
		9.2	0.561
		9.4	0.553
		9.6	0.544
		9.8	0.532
		10	0.518
$U(t) = e^{-iHt}$	6	Fit $S(t)$ for $t \in [0, 2.3]$	0.66
Obtained from slope in Fig. **??**	8	Fit $S(t)$ for $t \in [0, 4]$	0.63
	10	Fit $S(t)$ for $t \in [0, 5.8]$	0.61
	12	Fit $S(t)$ for $t \in [0, 7.6]$	0.60
$\Psi_A \otimes \Psi_B$	$6, 8, 10, 12$	3.2	1.24
		3.4	1.22
		3.6	1.19
		3.8	1.17
		4	1.16
		4.2	1.14
		4.5	1.08