Quantum Information Scrambling: Boulder Lectures

Brian Swingle[a]

[a] *Condensed Matter Theory Center, Maryland Center for Fundamental Physics, Joint Center for Quantum Information and Computer Science, and Department of Physics, University of Maryland, College Park, MD 20742, USA*

Abstract

These lecture notes introduce the physics of quantum information scrambling in local quantum systems, with a special focus on strongly interacting quantum matter and quantum gravity. The goals are to understand how to precisely quantify the spreading of quantum information and how causality emerges in complex quantum systems. In an attempt to clarify and strengthen some of the conjectured relationships between operator growth, information spreading, and entanglement growth, several new results are included in technical appendices and discussed in the main text. The lectures were originally delivered at the Quantum Information Boulder Summer School in Boulder, Colorado, July 2018.

Note: This document is a β version. In particular, many references are still missing, so check back soon. Please send any issues you notice to bswingle@umd.edu with subject "boulder notes". A permanent version with the same title will be made available on the arXiv at a later time. Last updated: July 25, 2018.

Survey

How does quantum information spread across a complex system? Are there fundamental bounds on quantum dynamics? How does locality emerge in quantum gravity? These are just a few of the questions that motivate the study of *quantum information scrambling.* These lectures are particularly focused on the way quantum information spreads in strongly interacting many-body systems. The ideas connect to many areas of physics, ranging from practical questions about how to reliably transmit quantum data to fundamental questions about how causality emerges from microscopic degrees of freedom in quantum gravity.

My particular goal in crafting these lectures is to explain two related issues. First, what precisely do we mean when we talk about quantum information

spreading and how do we quantify it? Second, what happens to information spreading in a strongly interacting system and how is this related to the emergence of causality in quantum gravity? At first glance, it may not seem that these topics are closely related, but I hope to convince you that they are.

Somewhat more precisely, I will argue that commutators of local operators provide a powerful way to diagnose quantum information propagation. These ideas are connected to recent developments in quantum chaos, particularly the special correlation functions known as out-of-time-order correlators (OTOCs) [1,2]. In my opinion, there is some lack of clarity in the field about what precisely OTOCs measure in local systems, so one goal of these notes is to convince you that they are tightly related to the spreading of quantum information, at least in one broad context. In fact, combined with some extensive appendices, I aim in the main text to construct a physicist's proof of this claim. (Again, for a broad, but not completely settled, class of systems.)

As part of the argument, I will explain how the butterfly velocity, defined from appropriate OTOCs, functions as a speed limit for quantum dynamics. You may be familiar with speed limits based on the speed of light or Lieb-Robinson bounds, in which case they can think of the butterfly velocity as a state-dependent, tighter version of such speed limits. In particular, the butterfly velocity places bounds on the expansion of entanglement in space and time. In attempting to clarify and strengthen some claims and conjectures in the literature, I have developed some results that are, to the best of my knowledge, more general than the existing analyses.

The notes are roughly structured as three lectures, with the corresponding original physical lectures being approximately one and a half hours each. I have a written a companion set of lecture notes that give a quantum information view of quantum chaos. They are slightly more introductory than these notes (which assume some familiarity with entanglement entropy, for example), and cover many related topics.

Comments are in gray. They are mostly meant to clarify notational issues, technical details, and other subtleties that lie outside the main line of development.

Exercise 0.1. *Exercises are in blue. These are mostly short calculations that flesh out the main discussion. They are meant to help you check your understanding of the discussion and to foreshadow subsequent developments.*

1 Introduction to quantum information propagation

Our interest here is the propagation and scrambling [3,4] of quantum information in local quantum many-body systems. This subject has numerous practical aspects, including the problem of sending quantum data from one place to another. It is also relevant in the study of quantum matter, where, for example, measures of information spreading have been related to electrical and thermal transport properties. More fundamentally, it is related to the emergence of locality and causality in quantum gravity. This may not mean much to you right now, but it should make a lot more sense by the end. These lectures will address all these topics, mostly focusing on information dynamics in strongly interacting quantum systems.

Let's start with information propagation in weakly interacting systems. Instead of trying to define this notion very carefully, I will just imagine that we are discussing the electromagnetic field, or some other system with freely propagating and weakly interacting waves or particles. For simplicity, consider a discrete one-dimensional system with some local degrees of freedom. The mathematical setup consists of a Hilbert space $\mathcal{V}$, a local Hamiltonian H, and some algebra of operators $\mathfrak{A}$. The Hilbert space is formed from the tensor product of local Hilbert spaces, $\mathcal{V} = \bigotimes_r \mathcal{V}_r$, which can be either finite dimensional (like spins) or infinite dimensional (like oscillators). The Hamiltonian is a sum of local terms with each term coupling together a finite number of nearby degrees of freedom. Modulo issues with infinite dimensional systems (which will we largely sidestep), the operator algebra $\mathfrak{A}$ can be taken to be the set of all operators.

To make things tangible, consider two parties, Alice and Bob, who are trying to communicate. Alice wants to send a bit $a \in \{0, 1\}$ to Bob given that they both have access to a shared physical system in state $|\psi\rangle$. Suppose that Alice and Bob each have access to some regions A and B of the system with A and B separated by a distance r. Depending on a, Alice applies a unitary U_a; alternatively, she can do nothing. Imagine she is exciting some wave designed to move over to Bob.

Figure 1: Mathematical setup in one dimension. Also shown are schematics of the regions that Alice (A) and Bob (B) have access to.

The system then evolves for a time t. Finally, Bob makes a measurement M_b of the system to try to learn Alice's bit. The probability that Bob obtains output $b = 0, 1$ given that Alice sent a is

$$\Pr(b|a) = \langle\psi|U_a^\dagger e^{\mathbf{i}Ht}M_b e^{-\mathbf{i}Ht}U_a|\psi\rangle. \tag{1.1}$$

An important requirement is that if Alice did nothing, then Bob should measure neither $b = 0$ or $b = 1$. This is the statement that

$$\langle\psi|e^{\mathbf{i}Ht}M_b e^{-\mathbf{i}Ht}|\psi\rangle = 0. \tag{1.2}$$

Subtracting this form of zero from $P(b|a)$, it follows that the probability to measure b given a is

$$\Pr(b|a) = \langle\psi|U_a^\dagger [M_b(t), U_a]|\psi\rangle, \tag{1.3}$$

where $M_b(t) = e^{\mathbf{i}Ht}M_b e^{-\mathbf{i}Ht}$ is the Heisenberg operator corresponding to M_b.

> If you like, think of Bob as having another measurement setting, $\emptyset$, corresponding to no signal from Alice. The measurement operators then obey $M_\emptyset + M_0 + M_1 = I$.

Let's pause to understand the physics of Eq. (1.3). First, at time zero, any operator $U_a \in \mathfrak{A}(A)$ and any operator $M_b \in \mathfrak{A}(B)$ commute provided A and B are disjoint, so information cannot yet be transmitted. More generally, if the commutator is non-zero but small, then the amount of information that can be transmitted is small, say as obtained from the classical Shannon formula for channel capacity. What does small mean? If the commutator is small as an operator, say in operator norm, then little information can be transmitted in any case; more generally, if at least the relevant matrix elements, say in state $|\psi\rangle$, are small, then again little information can be transferred. You may also

4

be familiar with the role of commutators in the theory of linear response where they represent the effect of a perturbation in the past. The main take away is that commutators are intimately connected with information propagation and causality.

Finally, while this discussion was framed in terms of classical information, the same analysis applies to the transmission of quantum data. Instead of sending a single classical bit, Alice can encode a qubit in some two level system, say the polarization of a photon or two hyperfine levels of an atom, and send the physical carrier to Bob. Presumably Bob would like to capture Alice's message and perform some operation on the quantum data it contains. However, this will also be unlikely to succeed if Bob's operators at time t approximately commute with Alice's operators.

Exercise 1.1. *Assume that Alice's message qubit is maximally entangled with a reference R that she continues to hold and that R and A are initially maximally entangled. Show that the entanglement between R and Bob's region B is small at time t if the commutator $[O_A, O_B(t)]$ is small for all O_A and O_B.*

There are many possible behaviors for the commutator of local operators in a quantum many-body system. For example, one might expect very different behavior between integrable and chaotic models, between non-interacting and strongly interacting models, and between localized and delocalized models. A natural first question is whether there are fundamental bounds on such commutators. You are probably familiar with at least one such constraint, namely the limitation on communication imposed by the speed of light. In the modern language of quantum field theory, this is called microcausality. It states that, given any two physical local operators $W(x)$ and $V(y)$ located at spacetime points x and y, their commutator must vanish if x and y are 'spacelike separated',

$$x, y \text{ spacelike separated} \rightarrow [W(x), V(y)] = 0, \qquad (1.4)$$

as illustrated in Fig. In other words, if y is outside of the 'light cone' of spacetime point x, then the corresponding operators must exactly commute. Crucially, this is an operator statement and hence a state-independent bound on information propagation. It is a fundamental property of any unitary Lorentz invariant local quantum field theory.

There is a somewhat analogous property for many lattice models which do not have relativistic causality built in microscopically. For discrete models with
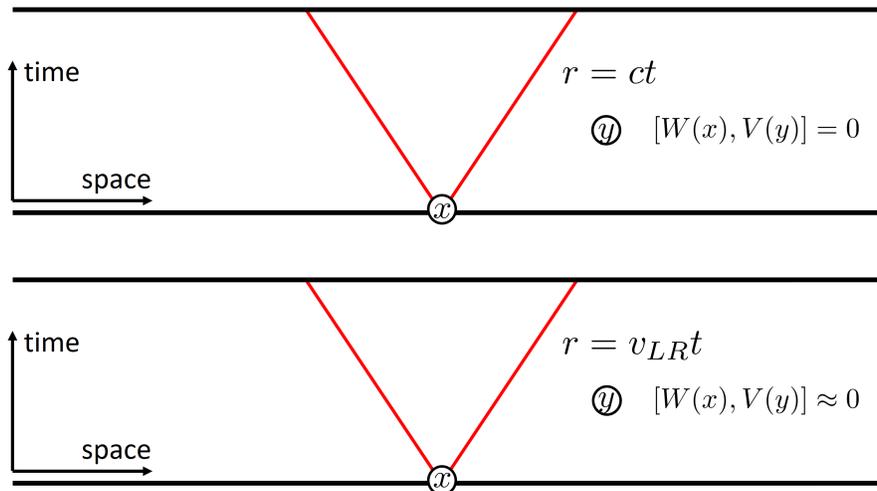
Figure 2: **Top:** Two points are spacelike separated if one lies outside of the light cone of the other. The commutator of local operators exactly vanishes for spacelike separated points. **Bottom:** The commutator of local operators approximately vanishes (as an operator statement) for points that are spacelike separated with respect to the Lieb-Robinson light cone.

a local Hamiltonian and a finite local Hilbert space dimension, it is possible to establish so-called Lieb-Robinson bounds that provide an upper bound on the size of commutators of local operators. For two local operators $W(r)$ and $V(r')$ at (spatial) positions r and r', one version of the Lieb-Robinson bound reads

$$\|[W(r,t),V(r',0)]\| \leq C\|W\|\|V\|e^{at-b|r-r'|} \tag{1.5}$$

where $\|\cdot\|$ denotes the operator norm. Observe that if $|r-r'| \gg \frac{at}{b}$, then the commutator is small as an operator since its norm is small. The ratio a/b is usually called the Lieb-Robinson velocity, $v_{LR} = a/b$, and it plays the role of the speed of light.

Any local operator $V(r')$ (with bounded norm) outside the 'Lieb-Robinson cone' of $W(r,t)$ will approximately commute with $W(r,t)$ up to small constant that decays at least exponentially with separation. By analogy with the speed of light constraint, we will also call this microcausality. I emphasize that this is a state-independent bound; it is universally applicable but possibly loose in many situations. As one minor point, one can show that the commutator decays faster than exponentially with $|r-r'|$ at very large $|r-r'|$. More physically, we will see further down in Sec. 1 and later in Sec. 2 that quantum information can propagate more slowly than what is required by a Lieb-Robinson bound. Furthermore, in Sec. 3 we will discuss the functional form of one measure of the size of commutators and Eq. (1.5) is not the general case.

6

We are particularly interested in strongly interacting quantum systems in these lectures. In such systems, the physics of excitations is typically very different than the nearly non-interacting waves of electromagnetism. In fact, such systems often cannot sustain any coherent excitation for very long, unless that excitation has a special reason for being protected, such as a sound mode or a Goldstone mode associated with a broken symmetry.

One precise manifestation of the lack of coherent excitations can be seen by looking at commutators (again). Given a system in thermal equilibrium at temperature $T = 1/\beta$ and given two non-conserved local operators $W(r)$ and $V(r')$ at positions r and r', one typically finds that

$$-\mathbf{i}\langle [W(r,t), V(r')]\rangle_\beta = -\mathbf{i}\mathrm{Tr}\left(\frac{e^{-\beta H}}{Z(\beta)}[W(r,t), V(r')]\right) \underset{t\gg\tau_r}{\approx} 0 \qquad (1.6)$$

regardless of the relative separation $|r - r'|$. Here τ_r is a relaxation time, something like the time it takes for the system to return to local thermal equilibrium after a perturbation. The intuition is that a system which is strongly interacting typically thermalizes, and a key characteristic of thermalization is the effective loss of memory of initial conditions. Since the expectation value of the commutator, with the factor of $\mathbf{i}$ to make it real, gives the change of $\langle W \rangle$ in response to a perturbation V in the past, it follows that if a system loses memory of initial condition after some time τ_r, then the corresponding commutator must be small. Shortly after a perturbation, measuring W gives no information about what happened in the past.

The non-conserved adjective is important because conserved quantities necessarily have much slower dynamics. For example, if W and V where taken to be the local density of some conserved quantity, say the the charge density, then the commutator has a form constrained by the fact that charge diffuses. Thus if the charge density is perturbed at some time, memory of that perturbation lasts much longer than for a typical non-conserved perturbation.

To get a visceral feel for these effects, let us consider a concrete model called the mixed field quantum Ising model. It is defined on a Hilbert space of n spin with the many-body Pauli operators, σ_r^α for $r = 1, \cdots, n$ and $\alpha = x, y, z$, obeying the algebra $\{\sigma_r^\alpha, \sigma_s^\beta\} = 2\delta^{\alpha\beta}\delta_{r,s}I_n$ and $[\sigma_r^\alpha, \sigma_s^\beta] = 2\mathbf{i}\epsilon^{\alpha\beta\gamma}\sigma_r^\gamma\delta_{r,s}$. The Hamiltonian, defined with open boundary conditions in one dimension, is

$$H = -J\sum_{r=1}^{n-1}\sigma_r^z\sigma_{r+1}^z - h_z\sum_r \sigma_r^z - h_x\sum_r \sigma_r^x. \qquad (1.7)$$

The first term assigns low energy to configurations with nearest neighbor spins aligned in the z basis. The second and third terms assign low energy to configurations with spins polarized along a particular axis. From the point of the σ_r^z

basis, the third term generates quantum fluctuations because it causes transitions between different σ_r^z eigenstates. It is conventional to set the nearest neighbor coupling J to unity, or in other words, to measure all energies in units of J and all times in units of $1/J$. The model then has two free parameters, h_x and h_z.

Exercise 1.2. *Find the lowest and highest energy states of the Ising Hamiltonian for $J = 0$ and for $h_x = h_z = 0$. Show that every product state in which each spin is in a state of definite y spin has zero average energy. Where do such states sit in the energy spectrum?*

There is an extensive literature on the physics of this Hamiltonian. For our purposes here, it will suffice to know that, for $h_z = 0$, the model is integrable, meaning in this case that it is secretly a model of non-interacting fermions in one dimension. On the other hand, the model with h_z non-zero is generically non-integrable and quantum chaotic and exhibits thermalization. This is visible in the expectation value of the commutators of local operators. In the free particle case, $h_z = 0$, there are long-lived excitations that can bounce around the system. On the other hand, in the non-integrable case, $h_z \neq 0$, there isn't any signature of a local perturbation that makes it very far in space or time.

Thus we have arrived at a new problem in the interacting case. It is not plausible that information has stopped spreading, but the expectation value of the commutator no longer seems to provide a very good way to measure information spreading. Said differently, it may be very hard to coherently transmit information as envisioned in our Alice-Bob exchange above. However, information is still spreading and we need a way to measure it. One clue, which will come up again later, is that while the diagonal matrix elements of the commutator might small, the commutator as an operator need not be. However, before explaining this in detail, let's look at a new way to track information spreading.

$\mathcal{B}$ $\qquad$ $\mathcal{B}$ $\qquad$ $\mathcal{B}$

Consider two orthogonal initial states, $|\psi_1\rangle$ and $|\psi_2\rangle$, which differ by the application of some local unitary operator W at site r_0:

$$|\psi_2\rangle = W|\psi_1\rangle. \tag{1.8}$$

In the Ising spin, the states could be $|\psi_1\rangle = |+y, \cdots, +y\rangle$ and $|\psi_2\rangle = |-y, +y, \cdots, +y\rangle$ with $W = \sigma_1^z$. Now introduce a new auxiliary system called the reference R which consists of a single qubit. The Hamiltonian dynamics does not act on the reference spin, which you may think of as sitting in an isolated box. Before the isolating the reference, however, prepare the entangled state

$$\frac{|\psi_1\rangle|0\rangle_R + |\psi_2\rangle|1\rangle_R}{\sqrt{2}}. \tag{1.9}$$

Given this initial entangled state, the idea is to track the entanglement with the reference as a function of time.

Let us suppose the states are chosen so that the reference is initially uncorrelated with the system away from r_0, so that only the site r_0 is entangled with R at time zero. This entanglement can diagnosed using the mutual information between R and site r_0. First define the von Neumann entropy of a set of spins A,

$$S(A) = -\text{Tr}(\rho_A \log_2 \rho_A). \tag{1.10}$$

Then the mutual information of A with B is

$$I(A : B) = S(A) + S(B) - S(AB), \tag{1.11}$$

where $S(AB)$ is the entropy of the union of A and B. For r_0 and R, a quick calculation gives

$$I(\{r_0\} : R) = 1 + 1 - 0 = 2. \tag{1.12}$$

This is the largest value that $I(\{r_0\} : R)$ can take, indicating maximal entanglement between r_0 and R.

> Note that if A and B are initially entangled, then even acting just on A conditioned on the state of R can lead to correlation between R and B.

> **Exercise 1.3.** *Show that for a pure state of ABR, the mutual information obeys $I(A : R) + I(B : R) = 2S(R)$ and that $I(A : R) \leq 2S(R)$. Use this to establish Eq. (1.12).*

Starting from the initially localized entanglement, you should expect the entanglement with the reference to expand out across the system in some fashion. One possibility is that the entanglement is carried in some coherent wavepacket throughout the system, remaining localized in space at any given time. This can occur under the right conditions in weakly interacting systems. But with strong interactions, the entanglement seems likely to spread and to become more complex. In other words, while at time zero the reference is entangled with a single spin in the chain, as time progresses, the reference will instead become entangled a complex collection of many spins.

This phenomenon of entanglement spreading is clearly visible in Fig. 3 which shows the dynamics of mutual information in the mixed field Ising chain with $n = 16$ spins. These data are obtained using a sparse matrix Schrodinger picture time evolution based on a Krylov method (which can be pushed to larger size if desired). Each line is the mutual information between the reference and an interval $[1, x]$ for various x as a function of time. As time passes, the information initially contained at the first site leaks more and more into the rest of the chain.
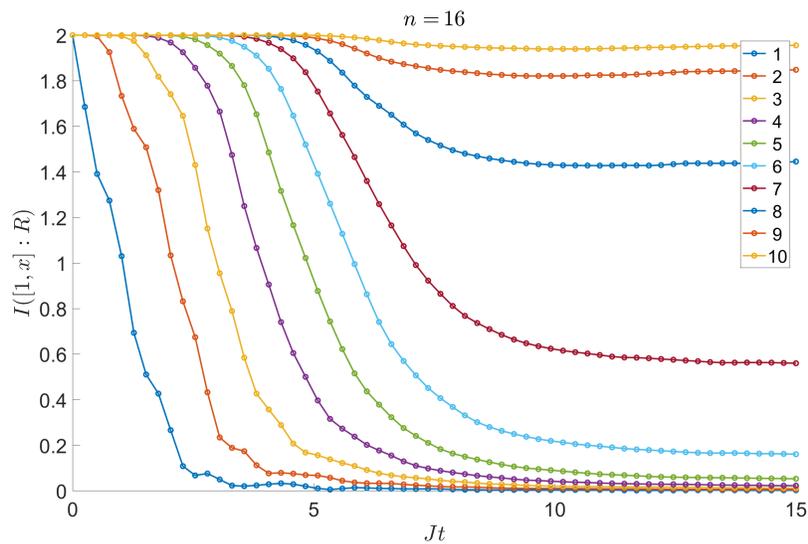
Figure 3: Mutual information between a reference qubit as subregions $[1, x]$ for $x = 1, \cdots, 10$ of an $n = 16$ spin chain. Notice how the mutual information always begins at 2, but then decays once the entanglement begins to leak out. However, if x is greater than half the system size, then the mutual information always remains above one. This is sketched in Fig. 4.
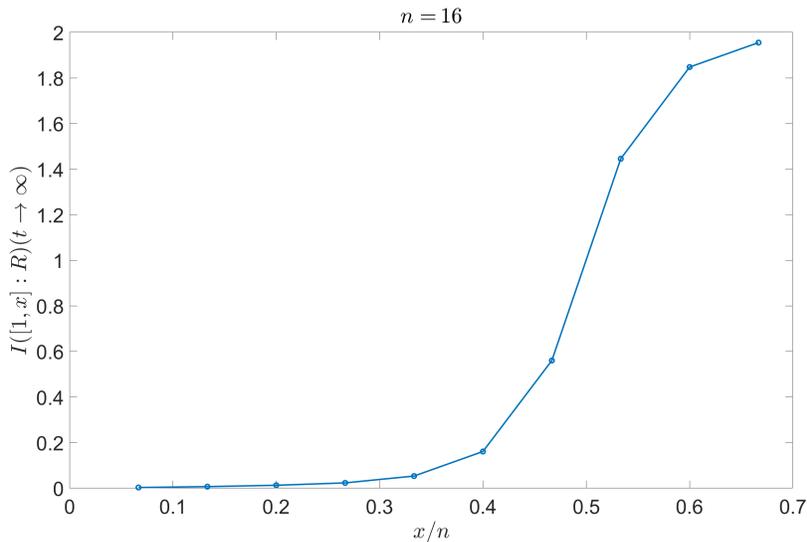
Figure 4: Late time value of the mutual information between a reference qubit as subregions $[1, x]$ for $x = 1, \cdots, 10$ of an $n = 16$ spin chain. Demonstrates the transition that happens in the late time value as x is increased past half the system size. Larger sizes show an even sharper transition.

Correspondingly, the entanglement is first recoverable given at least the first site, then given at least the first two, and so on. At late time, any region of more than half the spins is entangled with the reference as shown in Fig. 4.

> **Exercise 1.4.** *Given a tripartite pure state RAB with R a single qubit and $I(R : A) = 2$, show that there is a unitary on acting on A such one of the qubits of A is maximally entangled with R. Hint: Given the uncorrelated mixed state $\rho_{BR} = \rho_B \otimes \rho_R$, cook up two different purifications and use the fact that different purifications are related by a unitary on the purifier. For a challenge, work out the fidelity with a maximally entangled state between R and A in the case where $I(R : A) = 2 - \epsilon$ for small ϵ. Hint: Use Pinsker's inequality to bound the trace distance.*

As an aside, the approximate vanishing of the mutual information between the reference and any small subregion A of the spin chain is a consequence of thermalization. It means that no local operator in the small region A can distinguish the time evolved states $|\psi_1(t)\rangle$ and $|\psi_2(t)\rangle$. To see this, use the fact that $I(A : R)$ upper bounds connected correlations between R and A,

$$I(A : R) \geq \frac{\langle O_A O_R \rangle_c^2}{2\|O_A\|^2 \|O_B\|^2} = \frac{(\langle O_A O_R \rangle - \langle O_A \rangle \langle O_R \rangle)^2}{2\|O_A\|^2 \|O_B\|^2}. \tag{1.13}$$

11
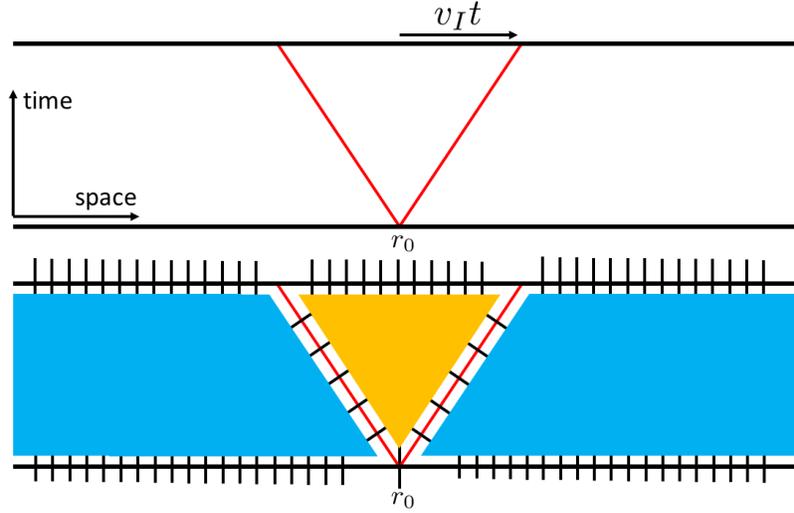
Figure 5: **Top:** Schematic of the lightcone structure with information velocity v_I. Note that the setup of the spin chain numerical calculation in Fig. 3 is slightly different, with the system consisting of a finite interval with r_0 at one end. **Bottom:** Approximate structure of the time evolution operator implied by the lightcone structure above. Information initially localized at r_0 is restricted to spread only by an amount $v_I t$. Note that, in practice, this will only be an approximate equality for the time evolution operator.

Assume that $I(A : R) \approx 0$ in the time-evolved version of Eq. (1.9). Then Eq. (1.13) with $O_R = \sigma_R^z$ implies

$$\langle \psi_1(t)|O_A|\psi_1(t)\rangle \approx \langle \psi_2(t)|O_A|\psi_2(t)\rangle. \tag{1.14}$$

Similarly, Eq. (1.13) with $O_R = \sigma_R^x \pm \mathbf{i}\sigma_R^y$ implies

$$\langle \psi_1(t)|O_A|\psi_2(t)\rangle \approx 0. \tag{1.15}$$

Together these conditions mean that observables in A cannot distinguish $|\psi_1(t)\rangle$ from $|\psi_2(t)\rangle$. Physically, this is the effective loss of memory of initial conditions associated with thermalization.

Exercise 1.5. *Prove Eq.* (1.13) *using Pinkser's inequality,* $S(\rho\|\sigma) \geq \frac{1}{2} Tr(|\rho - \sigma|^2)$, *where* $S(\rho\|\sigma) = Tr(\rho \log_2 \rho - \rho \log_2 \sigma)$ *is the relative entropy. Hint: the mutual information can be cast as a relative entropy.* *Factor of* $\ln 2$ *here?*

12

Perhaps the key feature visible in Fig. 3 is that the entanglement appears to spread ballistically. After a time t, we need access to a region of radius roughly $v_I t$ about the initial site to fully recover the entanglement. Here v_I is a kind of information velocity that we will try to constrain and understand. This intuition is depicted in Fig. 5 which shows at bottom a decomposition of the time evolution operator which would have the desired information propagation velocity.

Can we show that this decomposition of the time evolution operator is achievable? And what does all of this have to do with commutators, which we earlier used to diagnose information spreading? We already established that the expectation value of the commutator tends to decay rapidly in strongly interacting thermalizing systems. However, this decay does not imply that the commutator as an operator is small. Quite the opposite is true. To make sense of this, consider the next simplest object built from $[W(r, t), V(r')]$, the squared commutator,

$$C(r, r', t) = \langle [W(r, t), V(r')]^\dagger [W(r, t), V(r')] \rangle. \tag{1.16}$$

Intuitively, it is like a state-dependent operator norm of the commutator. Unlike the expectation value of the commutator, this object is positive definite and hence cannot suffer from the same kinds of cancellations that zeroed out the expectation value of the commutator. However, by the same token, this object is not so easy to interpret experimentally, for example, it is not obviously related to linear response. It is possible to measure it, given sufficient control over a system, but more on that later.

Exercise 1.6. *Given a one-dimensional quadratic boson model with momentum space creation $b_k^\dagger$ and annihilation b_k operators obeying $[b_k, b_q^\dagger] = \delta_{k,q}$ and Hamiltonian $H = \sum_k \epsilon_k b_k^\dagger b_k$, write an expression for the commutator $[b(x, t), b^\dagger(0, 0)]$ and hence the squared commutator. Does it depend on the state? Here $b(x) = \sum_k \frac{e^{ikx}}{L} b_k$ with L the system size.*

Assuming translation invariance for simplicity, $C(r, r', t)$ depends only on $r - r'$. We then introduce two new speeds, a lightcone speed v_L such that $C \approx 0$ for $|r - r'| > v_L t$ and a butterfly speed v_B such that $C = C_0$ for $|r - r'| = v_B t$ and some fixed order one constant C_0. In Sec. 3 we will understand the precise relationship between these two speeds, but for now we assume that they are equal (at least asymptotically, at large time). What is the physical behavior of $C(r - r', t)$ in the chaotic spin chain? As usual, as a function of time, it begins at zero since W and V initially commute, then once $|r - r'| \sim v_B t$ where v_B is the butterfly velocity, it rapidly increases in size until it saturates at late time. For unitary W and V the late time saturation value is two. What this tells us is that

13

while the expectation value is zero, the commutator itself is a highly non-trivial operator.

To complete this first section, I want to argue that $v_I = v_B$. Given a region A and a velocity v, define the region A_t to be A expanded in every direction by an amount vt. For example, if A is an interval $A = [a, b]$ than $A_t = [a - vt, b + vt]$. The precise claim is as follows. Start with a region A that is initially maximally entangled with a reference R. For any $v > v_B$, the entanglement with the reference can reliably recovered from A_t. By contrast, for any $v < v_B$, the entanglement can be reliably recovered from the complement $(A_t)^c$ of A_t. The key physical claim is that the butterfly velocity functions like a state-dependent information propagation velocity.

Recovery in A_t for $v > v_B$: The argument that information can be recovered in A_t for $v > v_B$ is reasonably intuitive. Roughly speaking, if commutators of local operators are all small, then the circuit should have a decomposition similar to that in Fig. 5 which implies that the entanglement can be recovered in A_t. More precisely, we can try to construct such a decomposition as follows. Separate H into two pieces, one localized in A_t, H_{A_t}, and the remainder, $\Delta H = H - H_{A_t}$. Now switch to the interaction picture to write

$$e^{-\mathbf{i}sH} = e^{-\mathbf{i}sH_{A_t}}\hat{U}(s) \tag{1.17}$$

where $\hat{U}(s)$ obeys

$$\mathbf{i}\partial_s\hat{U}(s) = e^{\mathbf{i}sH_{A_t}}\Delta H e^{-\mathbf{i}sH_{A_t}}\hat{U}(s). \tag{1.18}$$

The idea is that if $\hat{U}$ can be restricted to act non-trivially only on A^c, then we have the desired decomposition.

Such a restriction is almost certainly too strong as an operator statement, unless v is taken to be the microcausal limit, but what is reasonable is that $\hat{U}$, when acting on a suitable set of states Ω (for example, approximately uniform states of a certain energy density), can be taken to approximately act only on A^c. There are many pieces needed to make this argument precise, including establishing that H_{A_t} leads to similar commutators as H far from the boundary of A_t, arguing that not too much energy is produced, and so forth. The details are in App. C. The resulting argument is still not fully rigorous, although I think it could be made so. The basic idea is that if appropriate squared commutators are small, then as far as matrix elements within the set Ω are concerned, the operator $e^{\mathbf{i}sH_{A_t}}\Delta H e^{-\mathbf{i}sH_{A_t}}$ which generates $\hat{U}$ can be taken to have non-trivial support only on A^c.

Recovery in $(A_t)^c$ for $v < v_B$: The argument that the entanglement can be recovered only in $(A_t)^c$ for $v < v_B$ is more subtle. In particular, if v is only slightly smaller than v_B, then we still have access to most of the region that we just argued could be used to recover the entanglement. It turns out, however, that we really need all the region, for the same reason that we need more than half
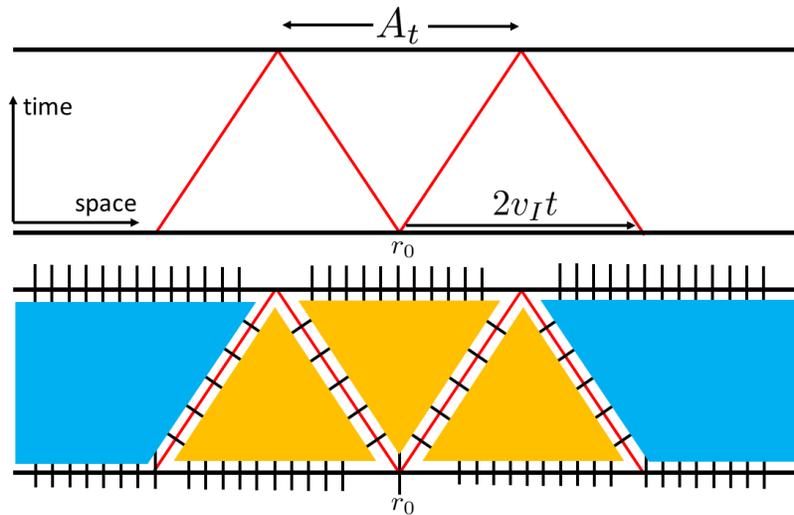
Figure 6: Given the ability to decompose a time evolution operator along a lightcone, say $v = v_B$, as in Fig. 5, we can always continue the construction. Supposing the initial state is a product state or some short-range entangled state, the decomposition shown here implies that, after removing the blue outer part of the time evolution, the region A_t is part of a pure state of size roughly twice the size of A_t. In other words, assuming a lightcone decomposition with $v = v_B$, then we would naturally need a region of radius at least $v_B t$ to have access to half the total pure state with which the reference is entangled.

the chain to recover the entanglement at late time. The intuition is sketched in Fig. 6 where we see that, assuming there is some optimal lightcone decomposition with $v = v_B$ (optimal meaning no smaller speed will do), the region A_t can be taken to be part of a pure state of roughly twice the size of A_t (assuming A itself is order one in size and we started with a short-range entangled state).

But how do we know that v_B is optimal? To see that, we turn to what is called the Hayden-Preskill protocol [3]. The idea, sketched in Fig. 7, is to consider some scrambling unitary in which one input is a qubit entangled with the reference R and the remaining inputs are entangled with a memory M. What the Hayden-Preskill analysis shows is that the entanglement with the reference can be recovered from just a few qubits of the scrambled output provided we also have access to the memory. The intuition, again, is that one needs to control more than half of the total system (output plus memory plus reference). The spin chain analog is shown in the right side of Fig. 7, where, assuming $v < v_B$, the memory is $(A_t^{(v_B)})^c$ and the small number of scrambled output bits is the difference between $A_t^{(v_B)}$ and $A_t^{(v)}$. The butterfly velocity appears here because it determines the size of the output which we can consider scrambled. Hence, the entanglement can be recovered from the memory plus the difference of $A_t^{(v_B)}$ and $A_t^{(v)}$ which taken together is $(A_t^{(v)})^c$. This is what we wanted to show, that the entanglement cannot be recovered in $A_t^{(v)}$ since it can be recovered in $(A_t^{(v)})^c$ (otherwise we violate no cloning).

You may still question whether having large squared commutator is really sufficient to indicate a fully scrambled output. In the context of Hayden-Preskill, this was argued for in Ref. [5]. Our setup is slightly different, for example, the entanglement between the input and the memory is not necessarily simply Bell pairs (it is generated by the lower orange triangles in Fig. 6), but presumably after distilling the entanglement into Bell pairs by acting on the complement of $A_t^{(v_B)}$ one could apply these standard results.

Before moving on, let's recap what happened. Starting from the problem of quantum information propagation, we showed how the commutator of local operators naturally arose as a measure of information spreading. However, in strongly interacting systems the expectation value of such commutators typically decays rapidly in time (or expands only slowly, as for a conserved density). Nevertheless, by studying the motion of entanglement we found there was still a sense in which information was spreading, at some definite speed. We then introduced the notion of squared commutators, which avoid the problem of a vanishing expectation value, and finally argued that the propagation speed set by the squared commutator was the same as the information propagation speed defined via entanglement. In this last analysis, the physics of quantum information scrambling played a crucial role in the argument.
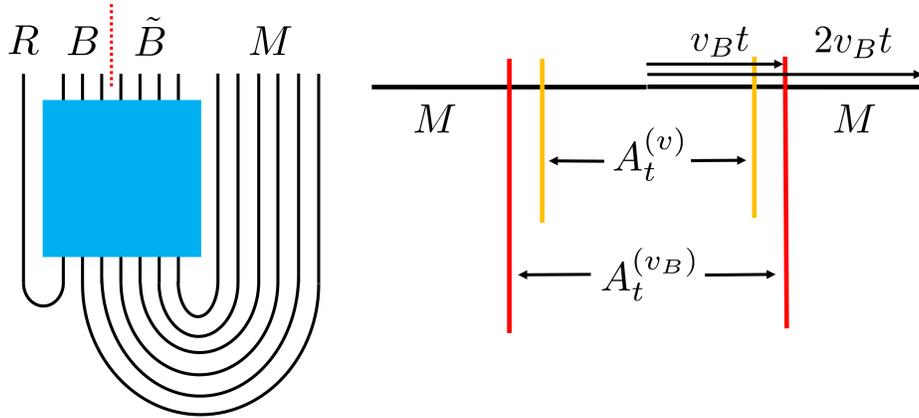
Figure 7: **Left:** Schematic of Hayden-Preskill thought experiment. The blue box is a unitary that scrambles its input while the input consists of a state that is maximally entangled with a reference R and a memory M (which are not acted on by the unitary). The result of the Hayden-Preskill analysis is that the entanglement with reference can be recovered from the memory plus a small number of additional output qubits B which may be a very small part of the total output. The intuition is that the memory M plus a small number of qubits B amounts to more than half the total system, so R is maximally entangled with it assuming the unitary scrambles. **Right:** An analog of Hayden-Preskill in the spin chain. Think of the region $A_t^{(v_B)}$ as the maximally scrambled output, while the memory is the part beyond $A_t^{(v_B)}$, the annulus from radius $v_B t$ to radius $2v_B t$. With the addition of the memory, the difference between $A_t^{(v)}$ and $A_t^{(v_B)}$ is the small number of scrambled output qubits needed recover the information. Hence the entanglement with the reference can be recovered from $(A_t^{(v)})^c$.

2 Gravity and information spreading

In this section we are going to revisit the discussion of Sec. 1 in the context of a particular model of quantum gravity known as the Anti de-Sitter space / conformal field theory (AdS/CFT) duality or holographic duality. This setting is interesting because it provides a highly non-trivial check of the previous arguments beyond lattice models and because it gives insight into the way causality emerges in quantum gravity.

We must begin with a breathtakingly brief review of the necessary bits of AdS/CFT. The basic statement of the duality is that some quantum system without gravity is equivalent to some seemingly different quantum system with gravity. The non-gravitational system is typically lower dimensional than the gravitational system, hence the duality is called holographic. The non-gravitational quantum system, the 'boundary', is associated with the boundary of the spacetime in which the gravitational quantum system, the 'bulk', lives. The boundary is typically a quantum field theory or some quantum many-body system. The bulk often arises from string theory, but in practice is typically described as a low energy effective theory coupled to gravity. There are many different versions of the duality, and we do not have a complete list of all quantum systems with gravity duals.

> Depending on how broadly we define the notion of a gravity dual, maybe every quantum system has one. However, one might want to reserve the moniker of gravity dual for special systems with a weakly coupled bulk. Or maybe any duality that involves gravity deserves the name, even if it looks nothing like gravity in our universe. In any event, these are probably not sharp divides.

To get into the details of the duality, we need to define some entries in the holographic dictionary which relates bulk and boundary. Think of the non-gravity side, the boundary, as a quantum field theory or as some lattice system describing a short-distance completion of a field theory. Think of the gravity side, the bulk, as a quantum field theory weakly coupled to Einstein gravity (meaning Newton's constant is small or that the Planck length is small, in appropriate units). This isn't the whole story for the bulk, but it is good enough for now. The first dictionary entry says that states on the non-gravity side map to states on the gravity side. The latter are described by a background geometry (the average value of the gravitational field, the metric) and a state of quantum fields coupled to gravity on that background. This is a kind of semi-classical limit where
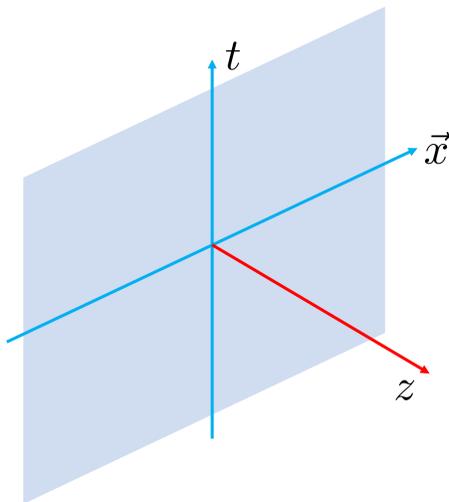
Figure 8: A schematic of AdS. The field theory directions are $\vec{x}$ and t (blue axes) while the radial coordinate is z (red axis). The asymptotic boundary of the spacetime, symbolized by the semi-transparent light blue plane, is located at $z = 0$.

the geometry is weakly fluctuating and we study small quantum perturbations (including gravitons, i.e., quantized metric perturbations) on top of the geometry. Note that this does not mean the geometry is non-dynamical, and different states can have very different geometries.

The simplest state to consider is the ground state of the boundary. For the case of a conformal field theory (CFT), the dual gravitational geometry is an almost empty spacetime suffused with a uniform negative energy density called the cosmological constant. This negative energy density causes the spacetime curve, and solving Einstein's equations, the desired solution is found to be Anti de Sitter space (AdS). How do we know which solution is the right one? Only AdS has the same symmetries as the ground state of a CFT, namely conformal symmetry. Incidentally, this is another dictionary entry, that symmetries must match on the two sides of the duality. Based on this, one could guess that AdS was dual to the ground state since it has all the right symmetries.

We will consider the AdS spacetime in the so-called planar limit, in which case it has a structure as sketched in Fig. 8. The 'field theory coordinates' label space $\vec{x}$ and time t. The 'radial coordinate' z labels the emergent direction. The metric in these coordinates reads

$$ds^2 = \frac{\ell^2}{z^2}\left(-dt^2 + d\vec{x}^2 + dz^2\right),\tag{2.1}$$

where ℓ is called the AdS radius and we have set the speed of light to unity. The boundary of the spacetime is $z = 0$ while $z \to \infty$ describes the deep interior. The

19

factor of $1/z^2$ is an example of what is called a warp factor (no relation to the Enterprise, unfortunately); it describes the way the gravitational field modifies the metric away from flatness.

> The very coordinate heavy approach we're using may annoy those with a relativist's bent. I sympathize, but unfortunately there isn't time do better.

The standard interpetation of the z coordinate is as length scale in the boundary. You may roughly imagine that the $z = \epsilon$ surface describes the microscopic field theory with some short distance cutoff, something like a lattice spacing or inverse momentum cutoff, of ϵ. Larger values of z then correspond to studying the field theory at a more coarse-grained scale. It is necessary to cut off the spacetime at some small $z = \epsilon$ to get finite answers to many questions, as we discuss below. The basic reason for believing the identification of z with length scale comes from the action of the conformal symmetry. The ground state of the boundary CFT is invariant under dilations which takes $\vec{x} \to \lambda \vec{x}$ and $t \to \lambda t$. In the bulk, this symmetry is realized as an invariance (isometry) of the metric where we also rescale $z \to \lambda z$.

> **Exercise 2.1.** *Verify that the AdS metric is invariant under the dilation of all coordinates by a factor of λ. Show that curves at fixed z have a length given by the usual Minkowski metric, up to a factor. Show that curves at fixed $\vec{x}$ and t have a length given by the $\ell \log \frac{z_2}{z_1}$.*

Another particularly important geometry is obtained from the dual of the thermal state in the boundary CFT. In the planar limit, the geometry is always an AdS-Schwarzchild black hole. If you have heard something about black holes having entropy, emitting radiation, and generally being good thermodynamic systems, this dictionary entry may seem reasonable. The black hole geometry, in the same coordinates, is

$$ds^2 = \frac{\ell^2}{z^2} \left(-f(z)dt^2 + d\vec{x}^2 + \frac{dz^2}{f(z)} \right) \tag{2.2}$$

where $f(z) = 1 - \left(\frac{z}{z_h} \right)^{d+1}$, d is the dimension of space in the CFT, and $z = z_h$ is the black hole horizon. We will discuss what happens behind the horizon later, so for now think of this geometry as just describing the exterior of the black hole. One of its key features is that time, as judged by an external observer far away, slows way down as the horizon is approached.

> **Exercise 2.2.** *Compute the proper distance from $z = \epsilon$ to the horizon at $z = z_h$. Compute the proper time it takes for a massive particle to fall from rest at $z = \epsilon$ to the horizon. At what coordinate time t does the particle reach*

� ✑ ✑

There are now many directions we could go in, including discussing correlation functions and clarifying the basic setup of the duality. However, in the name of brevity, I will discuss just two points, bulk causality and the entanglement-geometry connection. Bulk causality refers to the causal structure of the bulk metric. Given a parameterized curve $\{t(s), \vec{x}(s), z(s) | s \in [s_0, s_1]\}$, the tangent vector is

$$u^\alpha = \frac{dx^\alpha}{ds}, \tag{2.3}$$

where x^α runs over all the bulk coordinates. Given a metric $ds^2 = g_{\alpha\beta}dx^\alpha dx^\beta$ (with Einstein summation convention in force), a curve is classified as time-like, space-like, or light-like depending on whether $g_{\alpha\beta}u^\alpha u^\beta$ is less than zero, greater than zero, or equal to zero. The general rule is that causal trajectories must be time-like or light-like but not space-like, which would correspond to moving faster than the speed of light.

In the the AdS metric, Eq. (2.1), the causal structure at fixed z is the same as the flat Minkowski space in which the CFT lives. Presumably the speed of light limitation in the CFT directions follows from microcausality in the CFT. For an infalling particle not moving in the $\vec{x}$ directions, the causal limit is still $\left|\frac{dz}{dt}\right| < 1$, but it is not obvious what this corresponds to in the CFT. In the black hole metric, Eq. (2.2), even the causality in the CFT directions is modified by the factor $f(z)$. Near the asymptotic boundary at $z = 0$, where $f(z) \approx 1$, one recovers ordinary CFT causality, but near the black hole horizon causality seems mangled. One of the goals of this section is to explain how bulk causality arises using the ideas introduction in Sec. 1.

To do this, the other topic which must be discussed is entanglement. This set of ideas has another interesting history going back to Bekenstein's realization that black holes should have entropy. In AdS/CFT in the case of bulk Einstein gravity, there is a simple prescription known as the Ryu-Takayanagi (RT) formula to compute the entanglement entropy of any boundary region. The RT formula is illustrated in Fig. 9. Given a boundary region A, we consider all surfaces of the same dimension as A which terminate at $z = 0$ on the boundary ∂A of A. The entropy of A is the area in Planck units of the minimum area such surface,

$$S(A) = -\text{Tr}(\rho_A \log \rho_A) = \frac{|\gamma_A|}{4G_N} \tag{2.4}$$

where γ_A is the minimal surface and G_N is Newton's constant. Note that we use the terminology of 'surface' and 'area' no matter the dimension to speak about the general case.

The minimal surface depends on the background geometry. In the case of AdS with $d = 1$ boundary space dimensions, the minimal surface corresponding to a single boundary interval is a semi-circle. At non-zero temperature in the presence of the a black hole, small intervals have a minimal surface like the one in AdS, while large intervals have a minimal surface that hugs the horizon from the outside. This is actually crucial, since the entropy of a black hole is proportional to its surface area, so the RT surface running along the horizon recovers the black hole entropy.

Exercise 2.3. *Show that the minimal surface in AdS_{2+1} is a semi-circle and that its area (length) is proportional to $\log \frac{L_A}{\epsilon}$ where L_A is the length of $|A|$ and $z = \epsilon$ is a short distance cutoff. For a challenge, do the same thing for the black hole geometry in three bulk dimensions. It is possible to obtain an analytical formula.*

There is one more complication we have to add to the entanglement story. The RT formula describes only the leading order term in an expansion in Newton's constant. Remember that G_N was assumed small, so an entropy proportional to $1/G_N$ is large. The next term in the expansion is called bulk entropy and is given by the Faulkner-Lewkowycz-Maldacena (FLM) formula:

$$S(A) + \frac{|\gamma_A|}{4G_N} + S_{\text{bulk}}(\sigma_A) + \cdots . \tag{2.5}$$

The physical picture is that the bulk region σ_A contained between γ_A and A is associated to the boundary region A. Hence the entropy of A includes the leading area piece plus the von Neumann entropy of the bulk fields where the bulk is partitioned into σ_A and its bulk complement.

More precisely, the claim is that everything in region σ_A (see Fig. 9) can be represented or reconstructed in the boundary using only stuff in A. Hence if a particle in σ_A is entangled with another particle in the bulk complement of σ_A, then A should know about the entanglement. This is the physical origin of the second term, the FLM term, in the holographic entropy formula. The proposal that everything in σ_A is reconstructible in A is called entanglement wedge reconstruction. It has a covariant spacetime formulation valid for arbitrary geometries and regions A, but the simple spatial picture just outlined will suffice for now. One feature worth emphasizing is that everything in σ_A can be reconstructed in A and everything in the bulk complement of σ_A can be reconstructed in A^c, the boundary complement of A.

<center>✏   ✏   ✏</center>

We are now ready to get back to the story of quantum information propagation. First off, one can check (using a part of the holographic dictionary which
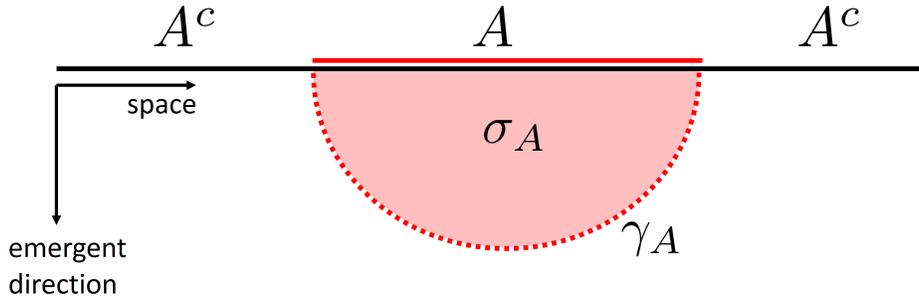
Figure 9: Ryu-Takayanagi prescription for the von Neumann entropy of a region A in the boundary, here shown for $d = 1$ boundary space dimensions. The time direction is suppressed. Given the region A, one finds the bulk curve γ_A of minimal length which terminates on ∂A at $z = 0$. The entropy of A (using the natural logarithm) is $S(A) = -\text{Tr}(\rho_A \log \rho_A) = \frac{|\gamma_A|}{4G_N}$ where $|\gamma_A|$ is the length of γ_A and G_N is Newton's constant. The enclosed region between A and γ_A is denoted σ_A. It represents the region of the bulk that can be reconstructed from the data in A alone.

I didn't explain) that, at non-zero temperature, the expectation values of commutators of local operators do indeed decay rapidly (excepting, again, conserved densities). The reason, basically, is that stuff falls into the black hole. However, we can still study the thought experiment where a local spin is entangled with a reference and the entanglement is allowed to expand. This process can actually be implemented very elegantly using AdS/CFT.

Start first with the ground state. The initial condition is a massive particle located near $z = 0$. The particle has two internal degrees of freedom, spin or some hyperfine state or whatever, and this internal degree of freedom is entangled with a reference (which, again, does not evolve in time). Starting from this initial condition, you know what happens to the particle from everyday experience: it falls down (towards the interior) due to the gravitational force. Assuming the particle is not moving in the $\vec{x}$ direction, it follows a time-like geodesic straight down as appropriate to a massive particle. As the particle accelerates, it moves closer and closer to the speed of light. For simplicity, take it to move at the speed of light on a trajectory $z = t$.

We use entanglement wedge reconstruction to track the entanglement with the reference. As long as the entanglement wedge of a boundary region B contains the infalling particle, the entanglement with the reference can be recovered from B. Conversely, if the infalling particle is contained in the entanglement wedge of the complement of B, then the entanglement with the reference can only be recovered from the complement. These claims are verified by calculating

$I(R:B)$.

The entropy of R is just one bit, since the reference is maximally mixed on its own. The entropy of B is given by the RT formula plus the FLM correction (divided by $\ln 2$ to convert to bits). If the infalling particle is in B, then since the entropy of B is unchanged by a unitary acting on the rest of the system, we may as well use that freedom to place the reference in the bulk in the entanglement wedge of B^c. Then the entropy of B is its value without the particle, $S(B)_0$, plus one bit for the bulk entanglement carried by the particle. If the infalling particle is not in B, the entropy of B is just $S(B)_0$. Similarly, for the entropy of BR, we may as well place the reference inside the entanglement wedge of B. If the infalling particle is in B, then the entropy is just the entropy of B without the particle since the bulk entanglement is contained entirely within B. If the infalling particle is not in B, then the entropy of B is $S(B)_0$ plus one bit for the bulk entanglement. Putting everything together, the result is

$$I(R:B) = 1 + [S(B)_0 + 1] - S(B)_0 = 2 \qquad \text{(particle in } B\text{)} \qquad (2.6)$$

$$I(R:B) = 1 + S(B)_0 - [S(B)_0 + 1] = 0 \quad \text{(particle not in } B\text{)}. \qquad (2.7)$$

Hence the question becomes how must B grow as a function of time such that the infalling particle is always contained in the entanglement wedge of B. Suppose B is a disk of radius r_B. The RT surface is given by specifying the radial coordinate z as a function of $r = |\vec{x}|$. A little geometry using Eq. 2.1 shows that the RT surface minimizes the area functional

$$\text{area} = \ell^d \int_0^{r_B} dr \frac{r^{d-1} \sqrt{1 + \left(\frac{dz}{dz}\right)^2}}{z^d}. \qquad (2.8)$$

Studying the corresponding Euler-Lagrange differential equation, one can show that $z(r) = \sqrt{z_{\max}^2 - r^2}$ is a solution for all d. In the RT surface equation, $z_{\max}$ is the maximum radial coordinate reached. Requiring that $z(r_B) = 0$ shows that $z_{\max} = r_B$. Hence if a particle is falling into the bulk at the bulk speed of light, then the radius of B must also expand at the boundary speed of light. A nice result indeed.

> **Exercise 2.4.** *Check that the semi-circle is indeed a solution to the Euler-Lagrange equations for Eq. (2.8) for any d.*

However, empty space or the corresponding ground state are just that, empty, so it is perhaps not surprising that the microcausal speed limit enters. On the CFT side, it is interesting to ask what happens when stuff is added to the system that might slow down motion like moving through a crowded room. On the gravity side, it is interesting to ask how bulk causality manifests in more complex geometries. These two questions are holographically dual to each other.

Consider the case of thermal equilibrium at temperature T. The bulk geometry is the AdS-Schwarzchild black hole of Eq. (2.2), and although the interpretation of bulk causality is less clear, the entangled infalling particle thought experiment can still be carried out. The geometry problems we have to solve are more complex, but an analysis is still possible as discussed in detail in App. D.

Two facts determine the result. First, infalling particles fall towards the black hole, accelerate to nearly the speed of light, and asymptotically approach exponentially close to the horizon at late time. Second, RT surfaces of large boundary regions stick close to the horizon for most of their extent. Combining these two pieces of physics, it is possible to show that, at large time t, a disk B whose entanglement wedge just contains the infalling particle must have a radius that increases with time like

$$r_B = \sqrt{\frac{d+1}{2d}} t + \cdots .$$ (2.9)

This growth of the entanglement wedge was first computed in Ref. [6], but without the particular interpretation we have given here.

Interestingly, this speed that is less than the speed of light for $d > 1$ boundary spatial dimensions. Moreover, as we anticipated in Sec. 1, it turns out to be identical to the butterfly velocity v_B obtained from OTOCs. I will explain a bit of the physics of OTOCs in holography below, so for now please just take this as a fact. The physical interpretation is that at temperature T, the CFT state consists of some hot strongly interacting relativistic plasma which tends to slow down the propagation of information.

In fact, App. D solves the aforementioned black hole geometry problems for a broad class of metrics known as hyperscaling violation metrics. We don't have time for the details, but one piece of physics that these geometries capture is a so-called dynamical exponent z (not to be confused with the radial coordinate) which allows for space and time to scale differently than in CFTs. For dynamical exponent z, space scales like $\vec{x} \to \lambda \vec{x}$ while time scales like $t \to \lambda^z t$. One still finds that v_B extracted from OTOCs gives the information propagation speed, but now

$$v_B \sim T^{1-1/z}$$ (2.10)

at temperature T. In particular, for $z > 1$ (which is required by microcausality), the butterfly velocity approaches zero as temperature is lowered. This indicates that quantum information can spread arbitrarily more slowly than any microscopic speed limit.

Exercise 2.5. *Obtain Eq. (2.10) using a scaling argument assuming that temperature scales like an inverse time and is the only scale in the problem.*

🍃 🍃 🍃

To explain the physical picture behind OTOCs in AdS/CFT, let us take a little detour to discuss the thermofield double state. This state is also interesting because it can be viewed as instantiating a particularly natural kind of reference system. Roughly speaking, the thermofield double is the minimal thermal bath that can produce thermal physics within a system. Thermality in a system coupled to an environment is typically understood as arising due to entanglement between the system and the environment. However, that entanglement is often complex and spread out (think infrared photons radiating from the Earth). The thermofield double is analogous to taking that complex environment and distilling in down to some kind of minimal form.

Here is the general construction. Given a system with Hamiltonian H and energy eigenstates $|E_i\rangle$, the canonical thermal state at temperature $T = 1/\beta$ is

$$\rho = \frac{e^{-\beta H}}{Z} \tag{2.11}$$

where $Z = \text{Tr}(e^{-\beta H})$ is the partition function. The thermofield double is a state defined on two copies of the system Hilbert space, referred to as the system and the purification. We will think of the systems as being placed side by side, left L and right R (meant to remind you of the reference), but not interacting. The thermofield double state is

$$|\text{TFD}\rangle = \sum_i \sqrt{\frac{e^{-\beta E_i/2}}{Z}} |E_i\rangle_L |E_i\rangle_R, \tag{2.12}$$

and its key property is that the state on just L or just R is exactly thermal. It is also possible to define a generalized time evolution with the left evolved for time t_L and the right for time t_R,

$$
\begin{aligned}
|\text{TFD}, t_L, t_R\rangle &= e^{-\mathrm{i}t_L H_L - \mathrm{i}t_R H_R} |\text{TFD}\rangle \\
&= \sum_i e^{-\mathrm{i}E_i(t_L+t_R)} \sqrt{\frac{e^{-\beta E_i/2}}{Z}} |E_i\rangle_L |E_i\rangle_R. \tag{2.13}
\end{aligned}
$$

Because of the special entangled structure of the state, this time evolved thermofield double only depends on the combination $t_L + t_R$.

Exercise 2.6. *Show that the infinite temperature limit of the thermofield double state, for a system of N qubits, is simply N Bell pairs connecting the two sides.*

In AdS/CFT, the thermofield double has a nice geometrical dual. If a single thermal state is dual to a black hole, then the thermofield double is dual
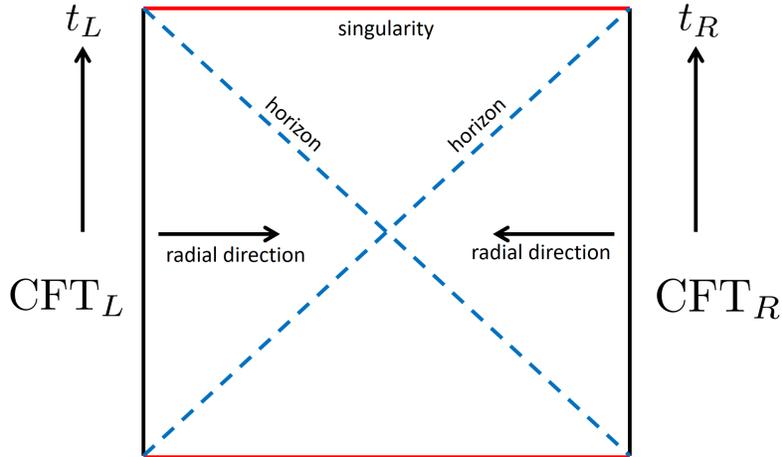
Figure 10: A schematic of the eternal AdS-Schwarzchild black hole space-time.

to two entangled black holes. As an example of the holographic connection between entanglement and geometry sketched above, these entangled black holes are represented geometrically by a wormhole which connects their interiors. The geometry is also called the eternal black hole geometry and can be obtained by looking for the maximal analytic extension of the single sided black hole geometry. It is depicted in Fig. 10, which shows a so-called Penrose diagram of the spacetime. This a spacetime picture that distorts lengths and times but preserves the causal structure so that the diagonal lines remain light-like.

Thinking the left L side as the original system, the previous discussion can be recast using the other side of the thermofield double as a souped up kind of reference which is entangled with the system. In this scenario, the entire entropy of the thermal density matrix is due to entanglement between the two sides of the thermofield double. Hence a thought experiment analogous to the entangled infalling particle problem can be instantiated using the thermofield double.

Start from the fact that there is a large amount of initial entanglement between two identical regions on either side of the thermofield double. At infinite temperature, where the thermofield double is simply many entangled pairs connecting to the two sides, this is particularly clear. As time evolves forward on the left side, the initially localized entanglement spreads in space. Correspondingly, the initial left-right mutual information between two identical regions decreases. However, given access to a region on the left with increasing size set by the butterfly velocity, the mutual information remains large. In fact, the mutual information can also be kept large by considering a region on the right of increasing size. Mathematically, Eq. (2.13) shows that time evolution on the left is equivalent to time evolution on the right. Physically, we use Hayden-Preskill

to recover the entanglement by expanding our access to the memory instead of expanding our access to the scrambled output qubits.

Using the tools and arguments discussed towards the end of Sec. 1, these conclusions can be generalized to a wider class of systems beyond AdS/CFT. In particular, the thermofield double setup provides what a 'thermal scale' generalization of the localized entanglement setup from Sec. 1. The key idea is that operators localized at the lattice scale typically change the energy of a state by an amount of order the microscopic couplings. However, at low temperature much below the microscopic scale, it is desirable to work with smeared operators which produce lower energy excitations when acting on a state, say excitations of energy of order the thermal energy T. Working with the thermofield double instantiates this smearing in particularly clean way and provides a simple tool to study information spreading at low energy.

$$\oslash \qquad \oslash \qquad \oslash$$

After familiarizing ourselves with the thermofield double state, let me explain how it is related to OTOCs. Imagine the following fanciful procedure: Evolve the left side, say, of the thermofield double back in time for a time $\Delta t_L = -t_W$. Now apply a simple operator W to the left side. Finally evolve the state forward in time again for a time $\Delta t_L = +t_W$. This returns the system to the original time, albeit with a small perturbation applied in the past. Suppose $t_W \gg \tau_r$, the relaxation time of the system. What is the effect of the perturbation from the distant past? From the point of view of either side alone, nothing much has happened. The right side remains literally untouched, while the left side has been perturbed, but the loss of memory due to thermalization suggests that the effects of the perturbation are not visible to simple probes.

If there is no discernible effect in of the perturbation in either side of the thermofield, how can we diagnose its presence? To get some intuition, let's think about the problem within the context of AdS/CFT. Suppose the perturbation in the past injected a localized packet of energy into the system, say a particle for simplicity. If the particle was near the boundary in the distant past, it will have fallen close to the black hole horizon by the present time. Such an infalling particle, having experienced a long period of acceleration due to the gravity of the black hole, will be moving close to the speed of light and will have enormous energy as reckoned by an observer hovering outside the black hole. Once strong enough, this energy source begins to significantly warp the geometry near the horizon as sketched in Fig. 11, leading to a wormhole which is effectively longer. Since correlations fall off with distance, this means that correlations between the two sides are weaker after the perturbation.
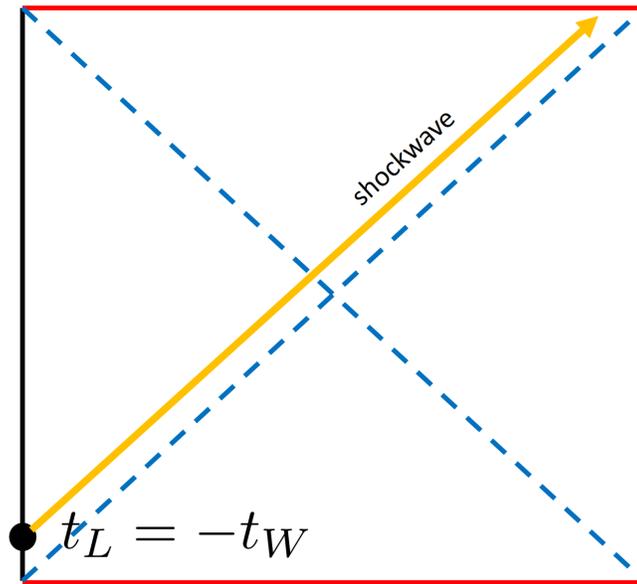
Figure 11: A schematic of the shockwave produced by an infalling particle moving near the speed of light. The shock disrupts the left-right correlations in the unperturbed thermofield double state. In the limit where the initial input energy is small and t_W is large, the shock hugs the horizon and Einstein's equations can be solved by patching together two halves of the eternal black hole geometry with a light-like shift across the shock.

A particle travelling at nearly the speed of light can be idealized as a light-like delta function of stress-energy pinned to the horizon of the black hole. This source creates what is known as a gravitational shockwave which expands outwards from the source. Passing through the shockwave has the effect of giving a kick in a certain light-like direction. This extra kick changes the geometry and leads to a decorrelation of the two sides of the perturbed thermofield double when the kick becomes strong. The strength of the kick goes like the proper energy, $E \sim e^{2\pi t_W/\beta}$, which is proportional to the boost the particle has experienced after falling for a time t_W. This exponential time dependence is one measure of quantum chaos in the system, with $\lambda_L = 2\pi/\beta$ being the now infamous quantum Lyapunov exponent of the black hole.

Since the perturbation effects correlations between the two sides of the thermofield double, let us understand these correlations more precisely. The perturbed thermofield double is

$$|W\rangle = e^{-\mathbf{i}H_L t_W} W_L e^{\mathbf{i}H_L t_W}|\text{TFD}\rangle, \tag{2.14}$$

where for any single sided operator O, $O_L = O \otimes I$ and $O_R = I \otimes O$. The correlation between V on the left and $V^{\mathbb{T}}$ ($\mathbb{T}$ is transpose in the energy basis) on the right is

$$\langle W|V_L \otimes (V^{\mathbb{T}})_R|W\rangle = \langle\text{TFD}|W_L^\dagger(-t_W)V_L W_L(-t_W) \otimes (V^{\mathbb{T}})_R|\text{TFD}\rangle. \tag{2.15}$$

To get intuition for this object, consider, say in a spin model with N spins, the limit $T = \infty$ where the thermofield double state is just a maximally entangled state $|\text{max}\rangle$ between the two sides. The state obeys a so-called operator pushing property whereby

$$O_R|\text{max}\rangle = (O^{\mathbb{T}})_L|\text{max}\rangle. \tag{2.16}$$

This allows us to move $V^{\mathbb{T}}$ from the right to the left. The two-sided correlation function then becomes a one-sided correlation function,

$$\langle W|V_L \otimes (V^{\mathbb{T}})_R|W\rangle_{T=\infty} = \frac{1}{2^N}\text{Tr}\left(W^\dagger(-t_W)VW(-t_W)V\right), \tag{2.17}$$

which is nothing but an infinite temperature OTOC at time $-t_W$. The negative time does not have any deep significance and is expected to lead to the same behavior as evolving W for positive time. In fact, if the behavior of the OTOC is generic for different local operators, then the previous claim follows from $\langle W^\dagger(-t)V_2W(-t)V\rangle_\beta = \langle W^\dagger V(t)WV(t)\rangle_\beta$, valid for any thermal state.

For general temperature T, the thermofield double still obeys an operator pushing property of a slightly more complex type,

$$O_R|\text{TFD}\rangle = (e^{\beta H/2}O^{\mathbb{T}}e^{-\beta H/2})_L|\text{TFD}\rangle, \tag{2.18}$$

where we interpret $e^{\beta H/2} O e^{-\beta H/2}$ as the Heisenberg operator $O(t)$ with imaginary time $t = -\mathbf{i}\beta/2$. Since imaginary time is periodic with period β, the circle represented by the coordinate $\tau = \mathbf{i}t$ is often called the thermal circle, and evolution by imaginary time $-\mathbf{i}\beta/2$ is called going half way around the thermal circle. The operator pushing property gives

$$\langle W|V_L \otimes (V^{\mathbb{T}})_R|W\rangle = \langle W^\dagger(-t_W)VW(-t_W)V(-\mathbf{i}\beta/2)\rangle_\beta \qquad (2.19)$$

or

$$\langle W|V_L \otimes (V^{\mathbb{T}})_R|W\rangle = \mathrm{Tr}\left(\sqrt{\rho}W^\dagger(-t_W)VW(-t_W)\sqrt{\rho}V\right). \qquad (2.20)$$

The right hand side of these expressions are called thermally regulated OTOCs (there are several varieties depending the positions of the operators on the thermal circle). They are generally believed to exhibit the same butterfly velocity as the non-regulated correlators, but this is not proven in general. At least the basic intuition still holds that if $W(-t_W)$ approximately commutes with V, then the thermally regulated OTOC collapses to something independent of t_W.

Exercise 2.7. *Verify Eq. (2.18).*

Now we see how to calculate thermally regulated OTOCs using AdS/CFT. Starting from the thermofield double state, we add a localized low energy perturbation in the past. This perturbation begins to strongly warp the geometry when it comes close to the horizon, leading to an expanding gravitational shockwave. Technically, one can solve Einstein's equations in the limit of a null delta function of stress energy hugging one horizon by patching together two halves of an eternal black hole geometry displaced by a relative light-like shift. Then measuring two-sided correlators in the shockwave geometry gives a thermally regulated OTOC. The butterfly velocity can then be extracted by studying the deviation of the OTOC from its initial early time value. The resulting velocity precisely matches the information propagation velocity obtained in Eq. (2.9).

🍃 🍃 🍃

Before moving on to the final section on the microphysics of operator growth and OTOCs, I want to tell you about one more important notion of entanglement growth. The setting is different but related to the considerations above. Consider a uniform initial with no special initial entanglement but which is out of equilibrium. For the mixed field Ising spin chain, this could be some initial product like the all $+y$ state. In AdS/CFT, one interesting class of examples is obtained by starting with empty space and making a uniform perturbation which increases the energy density of the system [7]. This kind of setup is called a global quench, because we start with a uniform out-of-equilibrium state and watch it

thermalize. In this context, one interesting question is how the entanglement entropy of a localized region increases with time.

Assuming thermalization takes place, the entropy of a region A much smaller than half the system size should look thermal at late time, $S(A) = s_T|A| + \cdots$, where s_T is the thermal entropy density at a temperature determined by the perturbed energy density. After a short initial period of local equilibration, but well before late time saturation, we expect the entanglement to grow at a constant rate proportional to the size of the boundary of A. This is because time evolution only increases the entropy of A thanks to terms in the Hamiltonian that act simultaneously on A and A^c. Extracting a factor of s_T and $|\partial A|$, the entanglement velocity v_E is defined during the period of linear entanglement growth by

$$\frac{dS(A)}{dt} = s_T|\partial A|v_E. \tag{2.21}$$

As defined, v_E seems analogous to a normalized entanglement growth rate, and although it has the units of a velocity, it is not clear what this velocity corresponds to physically.

As an aside, the thermofield double can also be used to obtain v_E. This may seem strange, since a single side is always in exact thermal equilibrium, but we know that the full state can exhibit time-dependence, as measured by two-sided correlations. The entanglement entropy of the combination of a region and its twin on the other side is time-dependent, and it exhibits similar phenomenology to the global quench entanglement growth. In particular, the entanglement velocity can be extracted from this setup. In AdS/CFT, this corresponds to studying the time-dependence of extremal surfaces stretching across the interior wormhole connecting the two black holes.

How is v_E related to v_B? I claim that $v_E \le v_B$, so that v_E obeys a causality bound. The statement that $v_E \le c$, with c the speed of light, has been proven in Lorentz invariant field theories in Refs. [8,9]. The stronger claim that $v_E \le v_B$ is plausible and has been conjectured, I am not aware of a careful argument in favor of it. In App. E I have tried to give a detailed argument for $v_E \le v_B$ roughly following the strategy of Ref. [9]. Here I will quickly mention the main idea, which is to compare, via the relative entropy, the out-of-equlibrium state of A and the thermal state restricted to A. The trick which makes this useful is using the decomposition shown in Fig. 5 to rewind the time evolution on a A.

The relative entropy is defined as

$$S(\rho\|\sigma) = \text{Tr}(\rho \log \rho - \rho \log \sigma). \tag{2.22}$$

Think of it as a way to compare two states, with $S(\rho\|\sigma) = 0$ if and only if $\rho = \sigma$. Key properties include positivity, unitary invariance, and monotonicity under partial trace. Given a time evolving state $\rho(t)$ and the thermal state σ with same energy density, we want to compare $\rho_{A_t}(t)$ and σ_{A_t} via relative entropy.

32

Here A_t denotes the expansion of A by an amount $v_B t$ and the time evolution operator U is assumed decompose as $U \approx U_{A_t} U_{A^c}$ (recall Fig. 5, but with A more than one site). The unitary invariance of relative entropy then yields

$$S(\rho_{A_t}(t)\|\sigma_{A_t}) = S(\text{Tr}_{(A_t)^c}(U_{A^c}\rho U_{A^c})\|\text{Tr}_{(A_t)^c}(U_{A^c}\sigma U_{A^c})). \qquad (2.23)$$

The right hand side obeys

$$S\left(\text{Tr}_{(A_t)^c}(U_{A^c}\rho U_{A^c})\|\text{Tr}_{(A_t)^c}(U_{A^c}\sigma U_{A^c})\right) \geq S(\rho_A\|\sigma_A), \qquad (2.24)$$

which follows from monotonicity of relative entropy under partial trace and the fact that U_{A^c} doesn't act on A.

This chain of reasoning gives the inequality

$$S(\rho_{A_t}(t)\|\sigma_{A_t}) \geq S(\rho_A\|\sigma_A). \qquad (2.25)$$

The physical interpretation is that everything that distinguishes ρ_A from σ_A also distinguishes $\rho_{A_t}(t)$ from σ_{A_t} since quantum information cannot spread by more than $v_B t$. Of course, $\rho_{A_t}(t)$ may be more distinguished from σ_{A_t} than ρ_A was from σ_A, hence the inequality. Assume that the reduction of the thermal state to A is approximately $\sigma_A \propto e^{-H_A/T}$ with H_A the Hamiltonian restricted to A up to boundary terms. Neglecting the boundary terms, the left hand side of Eq. 2.25 is

$$- S(\rho_{A_t}(t)) + S(\sigma_{A_t}) + \beta\langle H_{A_t}\rangle_{\rho_{A_t}(t)} - \beta\langle H_{A_t}\rangle_{\sigma_{A_t}}, \qquad (2.26)$$

and the right hand side is

$$- S(\rho_A) + S(\sigma_A) + \beta\langle H_A\rangle_{\rho_A} - \beta\langle H_A\rangle_{\sigma_A}. \qquad (2.27)$$

Recalling that ρ and σ have the same energy density, the Hamiltonian terms should cancel, again up to boundary terms that we are neglecting. The inequality Eq. (2.25) is then

$$S(\sigma_{A_t}) - S(\sigma_A) \geq S(\rho_{A_t}(t)) - S(\rho_A). \qquad (2.28)$$

Under the intuitively plausible situation that entanglement increases with sub-region size for regions less than half the system size, it follows that

$$|\partial A| s_T v_B t \geq S(\rho_A(t)) - S(\rho_A). \qquad (2.29)$$

The left hand side is the difference in thermal entropies between A and A_t (for $v_B t$ small compared to the linear size of A) while the right hand side is the total change in entanglement of A. The inequality $v_E \leq v_B$ then follows from the definition of v_E.

33

Following Grover, strong subadditivity actually implies that entanglement entropy is an increasing function of subsystem size up to half the system size for translation invariant systems. This is shown using $S(AB) + S(BC) \geq S(A) + S(C)$ (which follows from $S(AB) + S(AD) \geq S(A) + S(ABD)$ with D a purification of ABC). Working in $d = 1$ for simplicity, and setting $A = [0, x]$, $B = [x, x + \epsilon]$ and $C = [x + \epsilon, 2x + \epsilon]$, strong subadditivity gives $2S(x + \epsilon) \geq 2S(x)$.

3 Microscopic physics of operator growth

Having established that OTOCs measure the spreading of quantum information, in this section I discuss the microscopic physics of OTOCs in terms of the growth of Heisenberg operators. This discussion serves at least two functions. First, we will be able to understand the detailed spatial profile of the squared commutator $C(r, t)$ of local operators in generic chaotic systems. This has bearing on the question whether contours of constant C have different asymptotic velocities depending on the value of C. This turns out not the be the case, but there is a subtlety in that the ballistically expanding wavefront describing $C(r, t)$ is generically broadened. Second, we will understand better what the squared commutator and OTOCs have to do with quantum chaos in generic local systems.

As already explained in Sec. 1, the basic physical picture of $C(r, t)$ in a quantum chaotic system is that an initially simple local operator W, when evolved in time in the Heisenberg picture to $W(t)$, spreads in space and becomes more complex. However, it is not clear what the precise shape of the operator is, as diagnosed by $C(r, t)$. Let us first survey some representative examples from the literature.

We focus on the early growth region where C is still small. In the case of AdS/CFT, $C(r, t)$ takes the form [10]

$$C_{\text{holo}}(r, t) = \frac{1}{N_{\text{dof}}} e^{\lambda(t - r/v_B)} + \cdots . \tag{3.1}$$

Here N_{dof} is a measure of the number of onsite degrees of freedom, which is large in standard holographic models. Eq. (3.1) describes a ballistically travelling waveform with a sharp wavefront, meaning the shape of the waveform as a function of $t - r/v_B$ is independent of t. Many weak coupling calculations in field theory also find this form, at least near the wavefront. Next consider the case of a spin chain where the time evolution is taken to consist of alternating even and odd layers of random two qubit gates. In this random circuit model, the commutator ahead of the wavefront has the shape [11,12]

$$C_{\text{rand}}(r, t) \sim e^{-(r - v_B t)^2 / 4 D_B t} + \cdots . \tag{3.2}$$

In contrast to holography, this ballistically expanding wave has a diffusively broadened wavefront, meaning the scale over which C varies as a function of $u = r - t/v_B$ goes like $\sqrt{D_B t}$. Note that the random circuit model also has a version with a large number N_{dof} on each site, but while v_B and D_B depend on this number, but the holographic form is never obtained.

The question then arises as to which, if any, of these characteristic shapes describes the generic case with a finite local Hilbert space dimension. Unfortunately, this question cannot be reliably answered using small sized numerical simulations. These exhibit ballistic expansion with some kind of broadened wavefront, but it is not clear if the broadening will vanish in a large system or tend to the diffusive limit or have some other characteristic form. Non-interacting particles exhibit ballistic expansion of C with yet another characteristic broadening of the wavefront (C does not saturate at late time in this model, instead falling back to zero). You should not expect the non-interacting limit to be generic, but the spectre of multiple different universality classes is certainly raised.

There seems to be a lot of different possibilities, but it turns out to be possible to unify them all into a single framework. Shenglong Xu and I argued that the most general shape consistent ballistic operator growth can be characterized by one additional number, a broadening exponent p. The proposed general form, valid for large r and t with r/t greater than but close to v_B, is

$$C(r,t) = \exp\left(-\lambda \frac{(r - v_B t)^{1+p}}{v(vt)^p}\right). \tag{3.3}$$

This form is determined by demanding ballistic operator growth, an exponential decay with r at fixed r/t, and a finite logarithmic derivative $\frac{1}{C}\frac{dC}{dt}$ as $r, t \to \infty$. The AdS/CFT result fits the form with $p = 0$ (no broadening) while the random circuit result fits the form with $p = 1$ in $d = 1$ (diffusive broadening). Even the non-interacting fermion result fits with $p = 1/3$ in $d = 1$.

Exercise 3.1. *Verify that Eq. (3.3) obeys the three conditions listed just below the equation.*

I will try to convince you that the generic case in $d = 1$ is actually $p = 1$, corresponding to a diffusively broadened wavefront. As part of the story, I will explain how the $p = 0$ behavior is lost in a model that captures some of the features of the holographic case at large N_{dof}. One can make a corresponding analysis in higher dimensions, which is very interesting but is beyond what I can reasonably cover here.

Before moving on, however, I want to comment on the implications of Eq. (3.3) for the question raised above, whether different contours of constant C have different asymptotic velocities. Given the general shape in Eq. (3.3), the contours

obey

$$r_C = v_B t + \left(\frac{v(vt)^p}{\lambda} \log \frac{1}{C} \right)^{\frac{1}{1+p}}. \tag{3.4}$$

Hence, no matter the value of C, asymptotically one has

$$\lim_{t \to \infty} \frac{r_C}{t} = v_B. \tag{3.5}$$

However, at any finite t, the contour has an extra sub-ballistic time dependence going like $t^{\frac{p}{p+1}}$ which is due to the wavefront broadening. Thus at large time $C(vt,t)$ is very small for any $v > v_B$ and close to saturation (assuming chaos) for any $v < v_B$. We conclude that it is safe at large t to use v_B as both the scrambling velocity and the light cone velocity.

Exercise 3.2. *Derive Eq. (3.4) from Eq. (3.3).*

🍂   🍂   🍂

To gain a more microscopic understanding, it is useful to expand the Heisenberg operator $W(t)$ in a complete basis of operators. This expansion is exactly analogous to the expansion of quantum states in a complete basis for the Hilbert space. For a system of N qubits, the dimension of the Hilbert space of states is 2^N, and the dimension of the vector space of operators is 4^N. For a single qubit, a complete basis of operators is provided by the Pauli operators plus the identity, $\{I, \sigma^x, \sigma^y, \sigma^z\}$. For N qubits, a complete basis of operators is provided by Pauli strings $P_\mathcal{S}$. Each string is a tensor product of an operator from the set $\{I, \sigma^x, \sigma^y, \sigma^z\}$ for each qubit,

$$P_\mathcal{S} = \bigotimes_{r=1}^{N} \sigma_r^{\mathcal{S}_r}, \tag{3.6}$$

where $\mathcal{S} = (\mathcal{S}_1, \cdots, \mathcal{S}_N)$, $\mathcal{S}_r \in \{0,1,2,3\}$, and $\sigma^0 = I$.

Every operator can be expanded as

$$O = \sum_\mathcal{S} c(\mathcal{S}) P_\mathcal{S} \tag{3.7}$$

where the coefficients are

$$c(\mathcal{S}) = \frac{1}{2^N} \text{Tr}(O P_\mathcal{S}). \tag{3.8}$$

Note that the right hand side can be viewed as an inner product on operator space. The coefficients $c(\mathcal{S})$ are called operator amplitudes, while their absolute squares $|c(\mathcal{S})|^2$ are called operator probabilities. They are normalized to

$$\sum_\mathcal{S} |c(\mathcal{S})|^2 = \frac{1}{2^N} \text{Tr}(O^\dagger O). \tag{3.9}$$

36

Exercise 3.3. *Show that $(A, B) = \frac{1}{2^N} \, Tr(A^\dagger B)$ is an inner product on the vector space of operators.*

Let us see how to calculate the squared commutator using this formalism. For simplicity and because it will feature below, I will consider the infinite temperature state, but a general formula is straightforward to obtain. It is convenient that every Pauli string either commutes or anti-commutes with every other,

$$P_{\mathcal{S}_1} P_{\mathcal{S}_2} = q(\mathcal{S}_1, \mathcal{S}_2) P_{\mathcal{S}_2} P_{\mathcal{S}_1}, \tag{3.10}$$

where $q = \pm 1$. The infinite temperature OTOC of an operator O and Pauli string $P_\mathcal{S}$ is thus

$$F = \frac{1}{2^N} \mathrm{Tr} \left(O^\dagger P_\mathcal{S} O P_\mathcal{S} \right) = \sum_{\mathcal{S}'} |c(\mathcal{S}')|^2 q(\mathcal{S}, \mathcal{S}'). \tag{3.11}$$

🖉 🖉 🖉

The goal is now to analyze the operator amplitude dynamics in a representative model known as the Brownian coupled cluster model [13]. The model is useful because it is tractable and because it combines the relevant features of both AdS/CFT and the random circuit model. Like the random circuit model, it features a random time-dependent Hamiltonian, but unlike the random circuit model, it has an AdS/CFT-like large N_{dof} limit. Using it, we will be able to develop a physical picture why $p = 1$ is generic for one-dimensional chaotic systems.

The model can be defined in any dimension, but here we continue to focus on $d = 1$. The degrees of freedom are arranged in clusters which are then connected into a one-dimensional array. Every cluster contains N spin-1/2 degrees of freedom, and there are L clusters. The Hamiltonian is time-dependent and consists of two kinds of terms, within-cluster interactions and between-cluster interactions. To avoid mathematical complexities associated with stochastic calculus, it is simplest to present the model in discrete time.

The time evolution operator is

$$U(t) = \prod_{m=1}^{t/dt} \exp \left(-i \sum_r H_r^{(m)} - i \sum_{\langle rr' \rangle} H_{rr'}^{(m)} \right), \tag{3.12}$$

with m a discrete time index. The within-cluster terms and the between-cluster terms are

$$H_r^{(m)} = J_{m,r,a,b}^{\alpha\beta} \sigma_{r,a}^\alpha \sigma_{r,b}^\beta \tag{3.13}$$

$$H_{rr'}^{(m)} = g \tilde{J}_{m,r,r',a,b}^{\alpha\beta} \sigma_{r,a}^\alpha \sigma_{r',b}^\beta \tag{3.14}$$

where $\alpha, \beta \in \{0, 1, 2, 3\}$, $a, b = 1, \cdots, N$ label spins within a cluster, r, r' label clusters (sometimes called sites), and $\langle rr' \rangle$ means nearest neighbors. At each time step, the models contains two sets of uncorrelated random variables J and $\tilde{J}$ with mean zero and variance $\frac{1}{8(N-1)}dt$ and $\frac{1}{16N}dt$, respectively.

In the limit that $dt \to 0$, one can formulate a stochastic differential equation for the time evolution operator. From it, one can derive a master equation for the operator probabilities $\overline{|c(\mathcal{S})|^2}$ averaged over circuit realizations, i.e., over realizations of the couplings J and $\tilde{J}$. I will not get into the details of these equations here, but see Ref. [13] for complete details. The only important property we need is that $\overline{|c(\mathcal{S})|^2}$ depends only on the total number of non-identity Pauli operators on each cluster. This is technically an approximation, but it holds after a short time even if the initial condition does not obey it because the circuit average erases any distinction between the different Pauli operators. The total number of non-identity Pauli operators in $P_{\mathcal{S}}$ on cluster r is called the weight of the cluster and is denoted $w_r(\mathcal{S})$.

In this model it is convenient to analyze the operator averaged squared commutator,

$$C(r, t) = \frac{1}{3N} \sum_{a, \alpha \neq 0} \langle [W(t), \sigma_{r,a}^{\alpha}]^{\dagger} [W(t), \sigma_{r,a}^{\alpha}] \rangle_{T=\infty}. \tag{3.15}$$

Using the fact that the identity trivially commutes with $W(t)$, this can be rewritten using an operator averaged OTOC as

$$C(r, t) = \frac{8}{3}(1 - F(r, t)) \tag{3.16}$$

where

$$F(r, t) = \frac{1}{4N} \sum_{a, \alpha} \langle W(t) \sigma_{r,a}^{\alpha} W(t) \sigma_{r,a}^{\alpha} \rangle_{T=\infty}. \tag{3.17}$$

Using Eq. (3.11) with $W(t) = \sum_{\mathcal{S}} c(\mathcal{S}) P_{\mathcal{S}}$, it follows that the circuit-averaged OTOC is

$$\overline{F} = 1 - \frac{1}{N} \sum_{\mathcal{S}} w_r(\mathcal{S}) \overline{|c(\mathcal{S})|^2}. \tag{3.18}$$

For notational convenience, I will introduce $\phi(r, t)$ via

$$\overline{C(r, t)} = \frac{8}{3} \phi(r, t). \tag{3.19}$$

At early time, $\phi(r, t) \approx 0$ while at late time it saturates to $\phi(r, t) = \frac{3}{4}$, the fraction of non-identity Pauli operators.

Exercise 3.4. *Derive Eq. (3.18) using Eq. (3.11) and the stated assumptions.*

When N is small, it can be shown that $\phi(r, t)$ obeys a drift-diffusion equation as in the random circuit model. This leads to a circuit averaged $C(r, t)$ obeying

38

the universal form Eq. (3.3) with $p = 1$. Hence the Brownian coupled cluster model recovers the result of the random circuit model at small N.

At infinite N, something very different occurs. It can be shown that $\phi(r, t)$ obeys a so-called Fisher-Kolmogorov-Petrovksy-Piskunov (FKPP) type equation of the form

$$\partial_t \phi = (3 - 4\phi) \left(\frac{g^2}{2} \partial_r^2 \phi + (1 + g^2)\phi \right). \qquad (3.20)$$

Here g is the ratio of the strength of the between-cluster and within-cluster terms, and we have taken a continuum limit, which is qualitatively accurate. Although I will not explain its detailed derivation, one can see that this equation contains three essential pieces of physics: exponential growth in time, spreading in space, and saturation. The FKPP equation is very well known and describes a wide variety of physical processes including the propagation of combustion waves, the dynamics of invasive species, and the physics of certain quantum chromodynamics processes.

The key physical property of the FKPP equation is that, starting from a localized source, it supports travelling wave solutions with $\phi(r, t) = f(r - v_B t)$ where $v_B = \sqrt{18g^2(1 + g^2)}$ is the butterfly velocity. Well ahead of the front at $r = v_B t$, the waveform is

$$\phi(r, t) \sim e^{\lambda_L(t - r/v_B)}, \qquad (3.21)$$

which is Eq. (3.3) with $p = 0$. Hence the Brownian coupled cluster model also recovers the physics of AdS/CFT at large N. The exponent $\lambda_L = 6(1 + g^2)$ is an example of a quantum Lyapunov exponent.

> **Exercise 3.5.** *Obtain the butterfly velocity for FKPP by studying the linearized FKPP equation with a delta function source.*

🍃   🍃   🍃

Given the large and small N limits, the next question is how they are connected as N is varied. Physically, the infinite N limit functions to suppress quantum fluctuations, so that one may view the distribution $\overline{|c(\mathcal{S})|^2}$ as being concentrated on a single weight configuration. At finite N, quantum fluctuations occur, meaning that the distribution $\overline{|c(\mathcal{S})|^2}$ now assigns non-vanishing probability to different operators weight configurations. Let me emphasize that these fluctuations are proper quantum fluctuations. They are a consequence of the fact that $W(t)$ is a superposition of many different Pauli strings of different weight. In particular, the randomness associated with the couplings J and $\tilde{J}$ has already been averaged over and no longer enters the description. In essence, the circuit average serves to dephase the quantum operator amplitudes and convert

the Heisenberg equation of motion for the operator amplitudes into a master equation for the operator probabilities.

Following Ref. [13], I will call these quantum fluctuations 'noise'. In an abuse of notation where $\phi(r,t)$ now represents a noisy field, we obtain a noisy FKPP equation,

$$
\begin{aligned}
\partial_t \phi = {}& (3 - 4\phi)\left(\frac{g^2}{2}\partial_r^2 \phi + (1 + g^2)\phi\right) \\
& + \sqrt{\frac{1}{4N}(3 - 2\phi(r,t))\left(\frac{g^2}{2}\partial_r^2\phi(r,t) + (1 + g^2)\phi(r,t)\right)}\,\eta(r,t), \quad (3.22)
\end{aligned}
$$

where $\eta(r,t)$ is a white noise term representing quantum fluctuations. This noise term, while suppressed by $1/N$, has a dramatic effect on the physics. Notice also that it multiplicative noise, vanishing when $\phi = 0$, so it respects the causal structure.

The main effect of the noise term is to make the front position noise dependent. What this means is that the front continues to move with velocity v_B, but it is also randomly buffeted forward and backward as in a random walk. Within a particular noise realization, the wavefront is sharp and exhibits $p = 0$. However, the physical quantity in the quantum problem is the noise averaged value of ϕ. Close enough to the physical front at $r = v_B t$, the random walk nature of front position inevitably manifests and smears the sharp $p = 0$ front into a diffusive $p = 1$ front. Using the noisy FKPP literature, Ref. [13] showed that the corresponding diffusion constant was $D \sim \frac{1}{\log^3 N}$ at large N, a remarkably large value relative to standard $1/N$ corrections.

Given these developments, a conjecture and a corresponding physical picture naturally present themselves. I claim that, due to the inevitable presence of quantum fluctuations, generic one-dimensional quantum chaotic systems always have squared commutators obeying the universal form in Eq. (3.3) with $p = 1$. There is an analogous claim in higher dimensions, where the value of p depends on the dimension and is related to a random surface growth problem (the Kardar-Parisi-Zhang universality class). Based on what I have told you so far, the key pieces of evidence in favor of this claim are the random circuit model and the Brownian coupled cluster model. Interestingly, FKPP-like equations have also been obtained in a variety of large N and weak coupling calculations of squared commutators. These were all noiseless equations, but surely once quantum fluctuations are included, the dynamics will be governed by a FKPP-like equation with multiplicative noise and a corresponding broadened front.

The implications for quantum gravity are particularly interesting, and I hope to report on them soon. In short, it seems that the aforementioned quantum fluctuations lead, in the context of AdS/CFT, to a superposition of different gravitational shockwaves and a corresponding smearing of two-sided correlators

and two-sided entanglement. However, this is a preliminary conclusion that requires further analysis.

$$\oslash \qquad \oslash \qquad \oslash$$

Before wrapping up, I want to mention one final topic. I gave you physical picture and several pieces of evidence in favor of the claim that $p = 1$ is generic for local chaotic systems in one dimension. Still, it would be nice to check this claim in a quantum spin chain with no randomness in space or time and generic chaotic interactions. More generally, the preceding discussion did not provide a method to calculate squared commutators for generic physical systems. You may wonder if there is any classical method, analytical or numerical, which could treat generic systems since we are discussing a problem of quantum dynamics. Nevertheless, I want to briefly explain one recently developed tensor network method which can be used to access $C(r,t)$ in generic systems within a certain spacetime region [14].

It may surprise you to learn that a tensor network method can be useful here. Indeed, one typically expects that exact tensor network methods are restricted to early time in chaotic systems, since the dynamics produces entanglement rapidly. This is true in general, but it need not be the end of the story. The idea we will discuss involves switching from the Schrodinger picture to the Heisenberg picture. This is a useful switch because an initially local Heisenberg operator has an effective light cone. Outside the light cone, the operator is nearly the identity and so has limited operator entanglement. Hence it makes good sense to represent time-evolving Heisenberg operator $W(t)$ as a 'matrix product operator' (MPO), with the expectation that the representation will be accurate outside the light cone.

There are two important points to make about this approach. First, one can show rigorously, at least at infinite temperature, that if appropriate squared commutators are small beyond some distance r, then the corresponding Heisenberg operator has low entanglement across a cut located at r. Second, while an MPO with low entanglement can accurately capture the operator ahead of the front, it certainly cannot do so within the light cone. One should then worry about the possibility that truncation errors inside the lightcone might spoil the rest of the picture. However, the light cone saves us, since errors within the light cone should be unable to escape to spoil the dynamics outside. Hence we expect the MPO method will be valid for nearly the entire spacetime region outside the light cone, including the region where Eq. (3.3) is supposed to hold, so that both v_B and p can be extracted.

It is also worth pointing out that the time-evolving Heisenberg operator can be viewed as a state by acting it on a maximally mixed state or, more generally, on a thermofield double state. The operator entanglement discussed above is
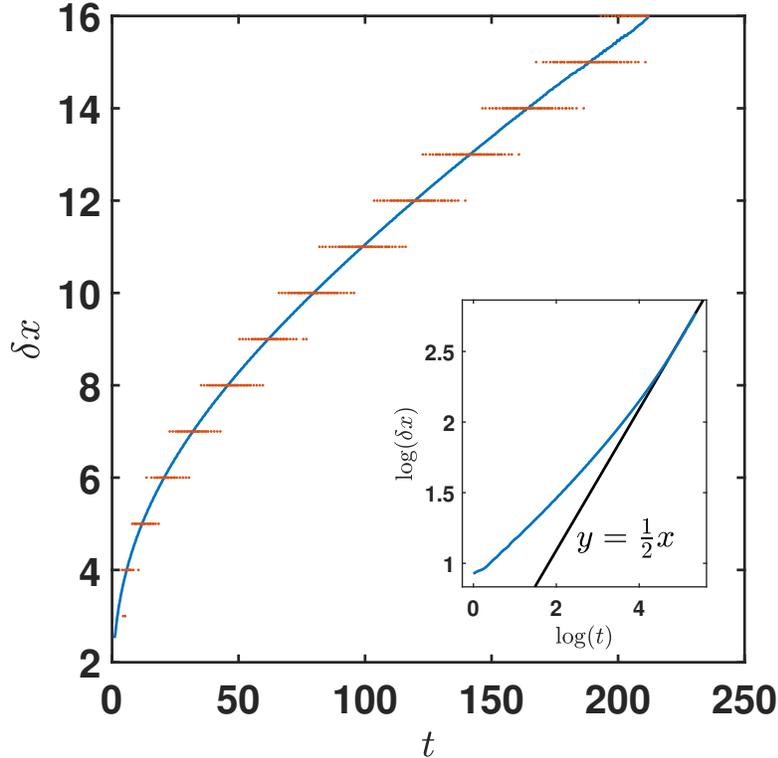
Figure 12: Separation between different contours of constant C as a function of time. The inset shows a log-log plot of the same data. The asymptotic approach to a slope of $1/2$, corresponding to $p = 1$, is clearly visible in the data.

just the entropy of the combination of a region and its twin on the other side of the thermofield double. In AdS/CFT, this construction yields the shockwave geometries considered in Sec. 2. Using those geometries, one can verify that at large N, the operator entanglement is small outside the light cone determined by the OTOC.

Fig. 12 shows data obtained from this MPO approach for the mixed field Ising model studied in Sec. 1 at infinite temperature. The simulation is performed for $n = 201$ spins out to quite long time, several hundred $1/J$. It has been checked that the results are converged in bond dimension with a bond dimension as low $\chi = 32$. What is plotted are the contours of constant squared commutator. The inset shows the difference between different contours as a function of time on a log-log plot. Eq. (3.4) predicts that the difference between contours should go like $t^{\frac{p}{p+1}}$, so on a log-log plot the data should approach a straight line of slope $\frac{p}{p+1}$. This is precisely what occurs with an asymptotic slope of $1/2$ corresponding to $p = 1$. Hence we verify that for a large, non-random chaotic spin chain, the

42

operator growth dynamics is ballistic with a diffusively broadened front exactly as predicted.

Epilogue

In these notes, I showed you to measure the spreading of quantum information using commutators of local operators. In generic chaotic systems, we found that information spreads ballistically, with a butterfly velocity v_B, and that v_B provides a speed limit to quantum dynamics, including entanglement growth. We showed that the butterfly velocity can be less than microscopic speed limits, sometimes dramatically so, and that this slowing down was essential for the emergence of causality in quantum gravity. Finally, we discussed the microscopic physics of operator growth, including a universal shape for squared commutators arising from quantum fluctuations and a tensor network method to study commutators in generic systems.

Still, this is all only the beginning. One crucial question is how to measure v_B in experiments. The simplest way is to directly measure $C(r,t)$, which is a hard but not impossible task. I and others have proposed a variety of experimental methods to measure OTOCs and squared commutators, all of which require some degree of quantum control often including the ability to effectively rewind time. Remarkably, several early experiments have already been carried out. So far, the experiments have probed regimes amenable to exact numerical simulation on a classical computer, but it seems plausible that the next generation of experiments will push beyond this threshold. Much more about these experiments can be found in my quantum chaos notes.

There many interesting directions for further theoretical work. One is information propagation in systems with long-range interactions. To the best of knowledge, it is still open whether power-law interactions always lead to super-ballistic information propagation. For example, existing Lieb-Robinson bounds cannot rule this out even for rapidly decaying power-law interactions. Other directions include proposed connections between chaos and transport of heat and charge and the interplay between the story here and the chaos bound at finite temperature [2] (Ref. [13] has a preliminary discussion). There is also much to understand about the breakdown of locality in quantum gravity.

In my opinion, we are only just starting to explore this exciting field of quantum information scrambling. With the many connections discovered so far and the prospect of new large-scale experiments on the horizon, I hope you consider getting into the field and bringing a new point of view.

References

[1] S. H. Shenker and D. Stanford, "Black holes and the butterfly effect,"
Journal of High Energy Physics (mar, 2014) 67, 1306.0622,
http://link.springer.com/10.1007/JHEP03(2014)067.

[2] J. Maldacena, S. H. Shenker, and D. Stanford, "A bound on chaos,"
Journal of High Energy Physics (aug, 2016) 106, 1503.01409,
http://link.springer.com/10.1007/JHEP08(2016)106.

[3] P. Hayden and J. Preskill, "Black holes as mirrors: quantum information
in random subsystems," *Journal of High Energy Physics* (sep, 2007)
120–120, 0708.4025, http://arxiv.org/abs/0708.4025
http://dx.doi.org/10.1088/1126-6708/2007/09/120
http://stacks.iop.org/1126-
6708/2007/i=09/a=120?key=crossref.eca9390b41dd1fa1b67fb6e27f9f17c0.

[4] Y. Sekino and L. Susskind, "Fast scramblers," *Journal of High Energy
Physics* (oct, 2008) 065–065, 0808.2096, http://arxiv.org/abs/0808.2096
http://dx.doi.org/10.1088/1126-6708/2008/10/065
http://stacks.iop.org/1126-
6708/2008/i=10/a=065?key=crossref.0eebed9ef6bfd908b2747d054e345a56.

[5] P. Hosur, X. L. Qi, D. A. Roberts, and B. Yoshida, "Chaos in quantum
channels," *Journal of High Energy Physics* (nov, 2016) 1–49, 1511.04021,
http://arxiv.org/abs/1511.04021
http://dx.doi.org/10.1007/JHEP02(2016)004.

[6] M. Mezei and D. Stanford, "On entanglement spreading in chaotic
systems," *Journal of High Energy Physics* **5** (May, 2017) 65, 1608.05101.

[7] H. Liu and S. J. Suh, "Entanglement Tsunami: Universal Scaling in
Holographic Thermalization," *Physical Review Letters* **112** (Jan., 2014)
011601, 1305.7244.

[8] H. Casini, H. Liu, and M. Mezei, "Spread of entanglement and causality,"
Journal of High Energy Physics **7** (July, 2016) 77, 1509.05044.

[9] T. Hartman and N. Afkhami-Jeddi, "Speed Limits for Entanglement,"
ArXiv e-prints (Dec., 2015) 1512.02695.

[10] D. A. Roberts, D. Stanford, and L. Susskind, "Localized shocks," *Journal
of High Energy Physics* **3** (Mar., 2015) 51, 1409.8180.

[11] A. Nahum, S. Vijay, and J. Haah, "Operator Spreading in Random
Unitary Circuits," *Physical Review X* **8** (apr, 2018) 021014, 1705.08975,
http://arxiv.org/abs/1705.08975
http://dx.doi.org/10.1103/PhysRevX.8.021014
https://link.aps.org/doi/10.1103/PhysRevX.8.021014.

[12] C. von Keyserlingk, T. Rakovszky, F. Pollmann, and S. Sondhi, "Operator hydrodynamics, OTOCs, and entanglement growth in systems without conservation laws," 1705.08910, http://arxiv.org/abs/1705.08910.

[13] S. Xu and B. Swingle, "Locality, Quantum Fluctuations, and Scrambling," *ArXiv e-prints* (May, 2018) 1805.05376.

[14] S. Xu and B. Swingle, "Accessing scrambling using matrix product operators," 1802.00801, http://arxiv.org/abs/1802.00801.

A   Further reading

- **Scrambling:**
- **Microcausality:**
- **Entanglement growth:**
- **AdS/CFT:**

B   Lieb-Robinson bound

This appendix reviews an elementary proof of a Lieb-Robinson bound for a simple one-dimensional spin to give a sense of how it works. The analysis follows a discussion of Osborne. Let's assume the Hamiltonian can be written as a sum of terms h_r that act on sites r and $r+1$. This can always be done coarse-graining any finite range interaction. Let the operator norm of h_r be J, which measures the local energy scale of the Hamiltonian.

Consider an operator W located at site r_0. The goal of Lieb-Robinson is to upper bound how far from r_0 this operator can spread after time t. The rough idea is to consider a series of approximations to $W(t)$ which involve truncating more and more distant terms in the Hamiltonian. These truncations then converge, roughly speaking, to $W(t)$ while also giving bound on the spreading.

Denote the restriction of H to the interval $[r-\ell, r+\ell]$ by H_ℓ, which means keeping only terms from H that are fully supported on the interval. The restricted Hamiltonian reads

$$H_\ell = \sum_{r=r_0-\ell+1}^{r_0+\ell-1} h_r \tag{2.1}$$

with the dependence on r_0 suppressed. Using the H_ℓ, define the sequence of Heisenberg operators W_ℓ via

$$W_\ell = e^{iH_\ell t} W e^{-iH_\ell t}. \tag{2.2}$$

To quantitatively estimate how these terms differ from each other, define the norms α_ℓ by

$$\alpha_\ell = \|W_\ell - W_{\ell-1}\| \tag{2.3}$$

with $\alpha_0 = \|W\|$. In terms of these, it is possible to upper bound objects of the form $\|W_\ell - W_{\ell'}\|$ as

$$\|W_\ell - W_{\ell'}\| \leq \sum_{j=\ell'+1}^{\ell} \alpha_j \tag{2.4}$$

by repeatedly adding and subtracting $W_{\ell''}$s and using the triangle inequality.

The α_ℓs are determined using a differential equation,

$$\frac{d}{dt}\alpha_\ell \leq \left\| \frac{d(W_\ell - W_{\ell-1})}{dt} \right\|. \tag{2.5}$$

Using the invariance of the norm under unitary transformations, the right hand side can be equivalently written as

$$\left\| \frac{d}{dt}(U_{\ell+1} W_\ell U_{\ell+1}^\dagger - W) \right\| \tag{2.6}$$

which is

$$\left\| [-iH_{\ell+1}, W_\ell] + [iH_\ell, W_\ell] \right\| = \left\| [H_{\ell+1} - H_\ell, W_\ell] \right\|. \tag{2.7}$$

The next step identifies $H_{\ell+1} - H_\ell$ with $h_{r_0+\ell} + h_{r_0-\ell}$, since these are the only new terms in $H_{\ell+1}$ fully supported on $[r_0 - \ell - 1, r_0 + \ell + 1]$ but not fully on $[r_0 - \ell, r_0 + \ell]$. We also use the fact that $W_{\ell-1}$ has no non-trivial support on $r_0 \pm \ell$ or $r_0 \pm (\ell + 1)$ and hence commutes with $H_{\ell+1} - H_\ell$. Thus the right hand side of the α_ℓ differential equation can be taken to be

$$\left\| [H_{\ell+1} - H_\ell, W_\ell - W_{\ell-1}] \right\| \leq 4J\|W_\ell - W_{\ell-1}\| = 4J\alpha_{\ell-1}. \tag{2.8}$$

Using $\|AB\| \leq \|A\|\|B\|$ and the triangle inequality, one has

$$\frac{d\alpha_\ell}{dt} \leq 4J\alpha_{\ell-1}. \tag{2.9}$$

The factor of four is a crude upper bound that takes into account both $h_{r_0+\ell}$ and $h_{r_0-\ell}$ which both appear twice due to the commutator.

Now we solve the upper limit of this system of differential equations with the initial condition that $\alpha_0 = \|W\|$ and $\alpha_{\ell>0}(t = 0) = 0$. The result is

$$\alpha_\ell(t) \leq \|W\| \frac{(4Jt)^\ell}{\ell!}. \tag{2.10}$$

This result is almost the end of the calculation. The remaining thing to do is to estimate the difference between W_ℓ and the true $W(t)$. This is

$$\|W(t) - W_\ell\| \leq \sum_{j=\ell+1}^{\infty} \alpha_j \leq \sum_{j=\ell+1}^{\infty} \|W\| \frac{(4Jt)^j}{j!}. \tag{2.11}$$

There are various ways to treat this infinite sum. For $\ell \gg 4Jt$, the simplest estimate is to say that it cannot by much larger than its first term, which is

quite small. More precisely, the ratio of term j to term $j + 1$ is $\frac{4Jt}{j+1} \leq \frac{4Jt}{\ell+2}$, so making even a crude approximation using a geometric series in $\frac{4Jt}{\ell+2}$ converges to something order one times the first term. After using Stirling's approximate for large ℓ, the first term is

$$\|W\|\frac{(4Jt)^{\ell+1}}{(\ell+1)!} \approx \|W\|\left(\frac{4eJt}{\ell+1}\right)^{\ell+1}. \tag{2.12}$$

This result corresponds to roughly the ℓth order in perturbation theory when expanding $W(t)$ in a Taylor series. Physically, it will describe the commutator dynamics for sufficiently small t and large ℓ.

Neglecting the difference between ℓ and $\ell+1$, the first term is order one when $\ell = 4eJt$. Setting $\ell_0 = 4eJt$, the first term can be written as

$$e^{\ell \log \frac{\ell_0}{\ell}}. \tag{2.13}$$

Using $-1 \geq -1 + \log \frac{\ell_0}{\tilde{\ell}}$ (valid for $\tilde{\ell} \geq \ell_0$) and integrating both sides from ℓ_0 to ℓ, it follows that

$$- (\ell - \ell_0) \geq \ell \log \frac{\ell_0}{\ell}. \tag{2.14}$$

The left hand side is the first order expansion of the right hand side in $\ell - \ell_0$, so the inequality states that going beyond first order only decreases the value. Hence

$$e^{\ell \log \frac{\ell_0}{\ell}} \leq e^{-(\ell-\ell_0)}, \tag{2.15}$$

or, using $\ell_0 = 4eJt$,

$$\|W(t) - W_\ell\| \leq \|W\|f(t)e^{4eJt-\ell}. \tag{2.16}$$

Here $f(t)$ is a polynomial prefactor that I wasn't careful about because it doesn't affect the basic exponential scaling. Note that the bound is definitely not tight at very large ℓ, since $1/\ell!$ decreases faster than $e^{-\ell}$. The bound is also trivial once $\ell < \ell_0$ because the right hand side is growing exponentially while the left hand side is bounded by $2\|W\|$.

Having established that W_ℓ is close to $W(t)$ for $\ell \gg Jt$, remains to upper bound the commutator. The idea is straightforward: If an operator V is a distance r from W, then an upper bound on the commutator $\|[W(t), V]\|$ is obtained by approximating W with $W_{\ell=r-1}$ since W_{r-1} exactly commutes with V. First add and subtract W_{r-1} inside the norm to give

$$\|[W(t), V]\| = \|[W(t) - W_{r-1} + W_{r-1}, V]\|, \tag{2.17}$$

and then use $[W_{r-1}, V] = 0$ and the bound on $\|W(t) - W_{r-1}\|$ to obtain

$$\|[W(t), V]\| \leq 2\|V\|\|W(t) - W_{r-1}\|. \tag{2.18}$$

Using the upper bound above, this is

$$\|[W(t), V]\| \leq 2\|V\|\|W\|f(t)e^{4eJt-r}, \tag{2.19}$$

which is Eq. (1.5) in the main text.

C Effective light cone velocity and squared commutators

This appendix gives one formalization of the idea of an effective light cone velocity v_L valid for a restricted set of states. The main idea is as follows. Given some region A and its complement A^c, the first step is to define a time evolved region A_t which is obtained by expanding every direction by an amount $v_L t + \ell_0$. The physical effect is encapsulated in the $v_L t$ term, which corresponds to information spreading at speed v_L. The ℓ_0 offset is technical device that simplies some of the analysis by giving us a little workspace. We then say that v_L is an approximate light cone velocity for some set of states if, for any region A, the time evolution operator U can be factorized as $U \approx U_{A_t} U_{A^c}$ as sketched in Fig. 5 and Fig. 6.

More precisely, consider a collection (not necessarily a linear space) of states Ω such that, for every state $|\psi\rangle \in \Omega$ the time evolution operator obeys

$$\|(U - U_{A_t} U_{A^c})|\psi\rangle\| < \epsilon \tag{3.1}$$

for all normalized states $|\psi\rangle$ of interest. As an example of the sort of set I have in mind, take all translation invariant states with a given energy density (within some small window that decreases with system size) plus their time evolved versions plus states obtained from these by acting with a finite number of local operator (including at different times). This kind of set is meant to describe states within a given energy window but where we rule out states with very unbalanced energy density by demanding translation invariance up to a small number of local perturbations which cannot add a macroscopic amount of energy.

Now let us see how this relates to the commutators. Take an initial state $|\psi\rangle$ and two local operators W and V such that $|\psi\rangle$, $W|\psi\rangle$, and $V|\psi\rangle$, and so forth, are all in Ω. For simplicity, take W and V to also be unitary, so that the squared commutator $C(t) = \langle [W(t), V]^\dagger [W(t), V] \rangle$ reduces to the OTOC,

$$F(t) = \langle W(t)^\dagger V^\dagger W(t) V \rangle. \tag{3.2}$$

If the supports of W and V are separated by more than $v_L t + \ell_0$, then there is region A such that V is contained in A and W is contained in $(A_t)^c$. Then it follows that

$$|F(t) - \langle \tilde{U}^\dagger W \tilde{U} V^\dagger \tilde{U}^\dagger W \tilde{U} V \rangle| < 4\epsilon, \tag{3.3}$$

where $\tilde{U} = U_{A_t} U_{A^c}$.

Now by assumption W is contained in $(A_t)^c$, hence $U_{A_t}^\dagger W U_{A_t} = W$. In the same way, V is contained in A, so $U_{A_c} V U_{A_c}^\dagger = V$. Note the similarity to the logic of the Lieb-Robinson bound. The second term in the absolute value above

is

$$\langle \tilde{U}^\dagger W^\dagger \tilde{U} V^\dagger \tilde{U}^\dagger W \tilde{U} V \rangle = \langle U_{A_c}^\dagger W^\dagger U_{A_c} V^\dagger U_{A_c}^\dagger W U_{A_c} V \rangle \tag{3.4}$$
$$= \langle U_{A_c}^\dagger W^\dagger V^\dagger W U_{A_c} V \rangle \tag{3.5}$$
$$= \langle U_{A_c}^\dagger V^\dagger U_{A_c} V \rangle \tag{3.6}$$
$$= 1. \tag{3.7}$$

Hence the actual OTOC obeys

$$|1 - F(t)| < 4\epsilon, \tag{3.8}$$

and the squared commutator obeys

$$C(t) = 2 - F - F^* \leq 8\epsilon. \tag{3.9}$$

It is natural to suspect that the implication goes in the opposite direction as well. We could hypothesize that if one had a suitable set of states such that OTOCs of simple local operators were small outside some effective butterfly cone $v_B t$, then the time evolution operator would factorize as above on these states. It is possible to demonstrate this conjecture with the assumption that OTOCs for all operators (not just local ones) on the appropriate regions are small, but it more challenging to make the argument using only OTOCs of local operators. Below I give a non-rigorous heuristic argument in favor of this conclusion.

C.1   Heuristic argument for $v_L = v_B$

Let's again work in one dimension for simplicity. The Hamiltonian is written $H = \sum_r h_r$. Given some region A and any speed $v > v_B$ (the butterfly speed from OTOCs), the idea is to approximate the time evolution operator U by $U \approx U_{A_t} U_{A^c}$ when acting on states in a set Ω. A_t is again the expansion of A by an amount $vt + \ell_0$ in every direction. To produce this decomposition, a natural idea is to break H into two pieces, with one piece, H_{A_t}, supported on A_t, and the remainder, $\Delta H = H - H_{A_t}$, supported on $(A_t)^c$. Going to the interaction picture with respect to H_{A_t}, the time evolution for time s (the target final time is t) is

$$U(s) = e^{-\mathbf{i}Hs} = e^{-\mathbf{i}sH_{A_t}} \hat{U}(s) \tag{3.10}$$

with

$$\mathbf{i}\partial_s \hat{U}(s) = e^{\mathbf{i}sH_{A_t}} \Delta H e^{-\mathbf{i}sH_{A_t}} \hat{U}(s). \tag{3.11}$$

As long as $e^{\mathbf{i}sH_{A_t}} \Delta H e^{-\mathbf{i}sH_{A_t}}$ can be well approximated by an operator acting on A^c, then the desired decomposition holds. One wants to show that since A is a distance $vt + \ell_0$ from $(A_t)^c$, the operator $e^{\mathbf{i}sH_{A_t}} \Delta H e^{-\mathbf{i}sH_{A_t}}$ is never appreciably supported on A.

Denote the H_{A_t} time evolution of ΔH by ΔH_{int},

$$\Delta H_{int} = e^{\mathbf{i}sH_{A_t}} \Delta H e^{-\mathbf{i}sH_{A_t}}. \tag{3.12}$$

Choose a site $r \in A$ and consider

$$\langle \psi_1 | q_r^\alpha \Delta H_{int} q_r^\alpha | \psi_2 \rangle - \langle \psi_1 | \Delta H_{int} | \psi_2 \rangle = \langle \psi_1 | q_r^\alpha [\Delta H_{int}, q_r^\alpha] | \psi_2 \rangle, \tag{3.13}$$

where $|\psi_i\rangle$ are states in Ω and $q_r^\alpha \in \{I, \sigma_r^x, \sigma_r^y, \sigma_r^z\}$. The absolute value of this quantity can be upper bounded by a squared commutator using Cauchy-Schwarz,

$$|\langle \psi_1 | q_r^\alpha [\Delta H_{int}, q_r^\alpha] | \psi_2 \rangle| \leq \sqrt{\langle \psi_2 | [\Delta H_{int}, q_r^\alpha]^\dagger [\Delta H_{int}, q_r^\alpha] | \psi_2 \rangle}. \tag{3.14}$$

Assuming that H_{A_t} leads to similar OTOCs to H (we will argue for this below), then since q^α is a simple local operator, it must be that its squared commutator with ΔH_{int} is small for any $|\psi_2\rangle \in \Omega$. Hence the matrix elements must approximately agree,

$$\langle \psi_1 | q_r^\alpha \Delta H_{int} q_r^\alpha | \psi_2 \rangle \approx \langle \psi_1 | \Delta H_{int} | \psi_2 \rangle. \tag{3.15}$$

To continue, expand ΔH_{int} in the q_r^α basis,

$$\Delta H_{int} = \sum_\alpha O^\beta q_r^\beta. \tag{3.16}$$

Introduce $c^{\alpha\beta}$ via

$$q_r^\alpha q_r^\beta = c^{\alpha\beta} q_r^\beta q_r^\alpha \quad \text{(no sum)}, \tag{3.17}$$

so that

$$\langle q_r^\alpha \Delta H_{int} q_r^\alpha \rangle = \sum_\beta c^{\alpha\beta} v^\beta \tag{3.18}$$

where

$$v^\beta = \langle \psi_1 | O^\beta | \psi_2 \rangle. \tag{3.19}$$

The constraint, using a matrix-vector notation, is

$$cv \approx v^0 \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}. \tag{3.20}$$

Given the vector identity

$$c^{-1} \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \tag{3.21}$$

it follows that $v^\alpha \approx v^0 \delta^{\alpha,0}$. This indicates that, at least as far as matrix elements within the set Ω are concerned, the $O^{\beta \neq 0}$ operators all approximately vanish. In other words, ΔH_{int} is approximately proportional to the identity for any site $r \in A$.

This result already makes it plausible that ΔH_{int} can be effectively restricted to A^c when acting on states in Ω. To argue for this more carefully, we would like to show that the action of ΔH_{int} is close to the action of $\mathcal{D}_A(\Delta H_{int})$ where $\mathcal{D}_A$ is the depolarizing channel on A which zeros out all non-identity operators in A. Working with qubits for simplicity, the action of the channel is

$$\mathcal{D}_A(\Delta H_{int}) = \frac{1}{4^{|A|}} \sum_{Q_A} Q_A \Delta H_{int} Q_A, \qquad (3.22)$$

where Q_A run over all Pauli strings in A which are identity on A^c. The difference is

$$\mathcal{D}_A(\Delta H_{int}) - \Delta H_{int} = \frac{1}{4^{|A|}} \sum_{Q_A} Q_A [\Delta H_{int}, Q_A], \qquad (3.23)$$

so one might hope to again bound the commutator using OTOCs as above. However, there is a subtlety because the Q_A need not be simple and can add considerable energy to the state, so we cannot guarantee that OTOCs between ΔH_{int} and Q_A cannot be guaranteed to be small.

An alternative approach is to argue as follows. Within A, ΔH_{int} can in principle contain both large Pauli strings with many non-identity Pauli operators and small Pauli strings with few non-identity Pauli operators. Because strings with many non-identity Pauli operators typically change the energy of a state by a lot, there matrix elements between states in Ω should be correspondingly small. Hence even if such operators are present, they can be dropped from ΔH_{int} in so far as matrix elements in Ω are concerned. It remains to rule out Pauli strings with a small number of non-identity Paulis. However, this is what our previous argument accomplishes, with the technical assumption that dropping the high weight Pauli operators from ΔH_{int} doesn't change the OTOC very much.

Since ΔH_{int} acting on states in Ω can be well approximated by the action of an operator with non-trivial support only in A^c, it follows that U also has the desired decomposition when acting on states in Ω. One applies the evolution equation for $\hat{U}$ to an arbitrary state in Ω, then one truncates the action of ΔH_{int} on the state to be non-trivial only in A^c. The desired operator is obtained by setting to zero the coefficients of all Pauli strings in the expansion of ΔH_{int} with non-zero weight in A.

C.2 Estimating energy production

It was implicitly assumed above that H_{A_t} leads to the same butterfly velocity as H, at least far from the boundary of A_t. This is reasonable precisely because of

the light cone, since it takes time for information from ∂A_t to propagate deep into A_t. However, one thing which is not clear is how much energy is produced by the truncation. More precisely, U exactly conserves energy, but $U_{A_t} U_{A^c}$ can in principle create energy when applied to some state. If it creates too much, then it could carry a state out of the set Ω which is undesirable. In particular, making a sharp cut in the Hamiltonian between H_{A_t} and ΔH is a very violent lattice scale operation.

The solution is to make the split between H_{A_t} and ΔH is a smooth way. Given a function $0 \leq f_r \leq 1$, consider two modified operators,

$$H_f = \sum_r h_r f_r \tag{3.24}$$

and

$$H_{1-f} = \sum_r h_r (1 - f_r). \tag{3.25}$$

If f is chosen to be unity deep inside A_t, zero far outside, and varying smoothly over a length scale ξ, then H_f is a smoothed version of H_{A_t} and H_{1-f} is a smoothed version of ΔH_{int}.

The next question is, given an infinitesimal time-step δt, how much energy does H_f add to the system. To gain some intuition, note that if f_r is uniform then H_f adds no energy. Hence the change in energy should be proportional to the gradient of f_r. The mathematical expression is

$$\delta \langle H \rangle = -\mathbf{i} \delta t \langle [H_f, H] \rangle. \tag{3.26}$$

Crucially, because energy is conserved locally, the commutator of h_r with H must be the lattice divergence of some energy current j^E. In one dimension this is $[h_r, H] = j^E_{r,r+1} - j^E_{r-1,r}$. This guarantees that if f_r is uniform, then

$$\sum_r f_r j^E_{r,r+1} - \sum_r f_r j^E_{r-1,r} = f \sum_r j^E_{r,r+1} - f \sum_r j^E_{r-1,r} = 0. \tag{3.27}$$

Since f_r is supposed to vary slowly, consider a continuum limit of the lattice formula,

$$\frac{\partial E}{\partial t} = \int d^d x \, f(x) \vec{\nabla} \cdot \langle \vec{j}^E \rangle (x), \tag{3.28}$$

where we generalized to d spatial dimensions. It still remains to determine $\langle \vec{j}^E \rangle$. The natural guess is that $\langle \vec{j}^E \rangle \propto \vec{\nabla} f$, since in thermal equilibrium the energy current should vanish. Therefore, we should have

$$\langle \vec{j}^E \rangle = \mu \vec{\nabla} f + \cdots, \tag{3.29}$$

where $\cdots$ vanish relative to the first term in the limit of constant f.

The energy production, for a slowly varying f, is thus

$$\frac{\partial E}{\partial t} = \mu \int d^d x \, (\vec{\nabla} f)^2. \tag{3.30}$$

Assuming the variation in f occurs on the scale ξ, the gradient is of order $1/\xi$ and the region of space in which the gradient does not vanish is of order $|\partial A_t| \xi$ (recall that f approaches one deep inside A_t. Hence the total energy produced is of order

$$\Delta E = \int dt \frac{\partial E}{\partial t} \sim \frac{\mu |\partial A_t| t}{\xi}. \tag{3.31}$$

In a chaotic system, the produced energy will diffuse away from its production location. It will spread over a length scale order $\sqrt{D_E t}$ with D_E the energy diffusion constant. Hence the final energy density is of order

$$\frac{\Delta E}{|\partial A_t| \sqrt{D_E t}} \sim \frac{\mu t}{\xi \sqrt{D_E t}}. \tag{3.32}$$

Hence if we take ξ to scale like $t^{1/2+\delta}$ for a target time t, then the final energy density is arbitrarily small at late time. Moreover, although such a ξ is quite large relative to microscopic scales, it can always be taken to be a small correction relative to $v_L t$ (for the purposes fitting the slowly varying ξ scale into the expanded part of A_t).

D   Growth of entanglement wedge

This appendix contains the details of the growth of the entanglement wedge calculation. This calculation was already carried out for AdS-Schwarzchild black holes in Ref. [6] where they observed that the rate of growth of the entanglement wedge matched the butterfly velocity as obtained from OTOCs. In the main text we argued in Sec. 1 that this had to be case. Here we check this claim for a much wider class of the bulk geometries known as hyperscaling violation black holes. From a physical point of view, this class of geometries is interesting because it includes the possibility of a different scaling of space and time and includes phenomena that bear some resemblance to more conventional materials with a finite density of electrons.

The hyperscaling violation metric in the planar limit is

$$ds^2 = \frac{\ell^2}{u^2} \left(\frac{u}{u_0} \right)^{2\theta/d} \left(-f(u) \frac{dt^2}{(u/u_0)^{2z-2}} + \frac{du^2}{f(u)} + d\vec{x}^2 \right) \tag{4.1}$$

where $f(u) = 1 - (u/u_h)^{d-\theta+z}$. The constant u_0 is a length that ensures the various quantities appearing in the metric have their engineering dimensions. We have replaced $z \to u$ relative to the coordinates in the main text. The symbol z in this appendix denotes the dynamical exponent, which relates the

scaling of length and time via $\vec{x} \to \lambda\vec{x}$ and $t \to \lambda^z t$. The parameter θ is called the hyperscaling violation exponent. It measures the deviation of the scaling of the entropy density relative to the CFT value. Note that the limit $z \to 1$ and $\theta \to 0$ recovers the AdS-Schwarzchild metric, Eq. (2.2), provided we identify u with the z-coordinate there.

As in the main text, a particle released near boundary will fall towards the black hole and quickly approach the near horizon region. In this region, where the particle asymptotically approaches the horizon, it is possible to perform a analysis that is independent of much of the details of the problem. To carry out the analysis near the horizon, it is useful to rewrite the metric in terms of the proper radial distance to the horizon. The proper distance ρ is defined by

$$d\rho = -\frac{\ell}{u}\frac{u^{\theta/d}}{u_0^{\theta/d}}\frac{du}{\sqrt{f}}. \tag{4.2}$$

Expanding $f(u)$ near the horizon gives

$$f \approx (d - \theta + z)(u_h - u)/u_h, \tag{4.3}$$

so the proper distance near the horizon is

$$\rho = \ell\frac{u_h^{\theta/d}}{u_0^{\theta/d}}\frac{2\sqrt{1 - u/u_h}}{\sqrt{d - \theta + z}}. \tag{4.4}$$

In terms of ρ, the near horizon metric is

$$ds^2 = -\left(\frac{d - \theta + z}{2}\frac{u_0^{z-1}}{u_h^z}\right)^2 \rho^2 dt^2 + d\rho^2 + \frac{\ell^2}{u_h^2}\left(\frac{u_h}{u_0}\right)^{2\theta/d}d\vec{x}^2. \tag{4.5}$$

Analytically continuing $t \to -i\tau$, periodicity in imaginary time, $\tau \sim \tau + \beta$, makes the geometry a cone. Demanding regularity of the cone at $\rho = 0$, i.e., requiring that if the metric is $d\rho^2 + \rho^2 d\theta^2$, then θ must be periodic with period 2π, fixes the inverse temperature to be

$$\beta = \frac{4\pi}{d - \theta + z}\frac{u_h^z}{u_0^{z-1}}. \tag{4.6}$$

The next step is to determine the near horizon geodesics. Defining the coordinate η via $t = \frac{\beta}{2\pi}\eta$ and suppressing the transverse directions, the near horizon metric takes the simple form

$$ds^2 = -\rho^2 d\eta^2 + d\rho^2. \tag{4.7}$$

You may recognize this as Rindler space, the part of a flat spacetime associated with a uniformly accelerating observer. All these coordinate transformations were useful because the geodesics in this geometry are simple (after one more

coordinate transformation). Since Rindler space is just a part of flat Minkowski space, one can determine geodesics in it by changing variables to Cartesian coordinates, $X = \rho\cosh\eta$ and $T = \rho\sinh\eta$. In X, T coordinates, geodesics are straight lines, $X = X_0 + VT$. In the Rindler coordinates this is

$$\rho = \frac{X_0}{\cosh\eta - V\sinh\eta}. \tag{4.8}$$

Hence, at late times when $\eta = 2\pi t/\beta$ is large, the trajectory obeys

$$\rho(t) \approx \frac{2X_0}{1-V}e^{-2\pi t/\beta}. \tag{4.9}$$

This means that, according to Eq. (4.4), the infalling particle gets exponentially close to the horizon at late times.

The second part of the calculation is to determine the minimum size of a boundary region whose entanglement wedge just contains $\rho(t)$. It is easy enough to solve this problem in special cases numerically, but a general analysis is possible when a boundary region A is large compared to the thermal scale. In this case, the RT surface of A hugs the horizon for much of its extent, and the essential physics can be understood from the shape of the RT surface in the near horizon region.

Let A be a d-dimensional disk of radius r_A. We may specify the RT surface by giving ρ as a function of $\vec{x}$ in A. Because the RT surface hugs the horizon for a large part of its extent, it is useful to expand the area functional to quadratic order in ρ in the near horizon region. The full near horizon spatial metric, to second order in ρ, is

$$ds^2 = d\rho^2 + \frac{\ell^2}{u_h^2}\left(\frac{u_h}{u_0}\right)^{2\theta/d}\left(1 + m^2\rho^2\right)d\vec{x}^2 \tag{4.10}$$

where m^2 is

$$m^2 = \frac{2(d-\theta)}{d}\frac{(d-\theta+z)u_0^{2\theta/d}}{4u_h^{2\theta/d}\ell^2}. \tag{4.11}$$

Changing variables to $y^i = \frac{\ell u_h^{\theta/d}}{u_h u_0^{\theta/d}}x^i$, the area is

$$\text{area} = \int d^d y \det\left(\partial_i\rho\partial_j\rho + \left(1 + m^2\rho^2\right)\delta_{i,j}\right). \tag{4.12}$$

Using $\det(1+X) \approx 1 + \text{tr}(X)$ valid for small X, the small ρ expansion reads

$$\text{area} \approx \int d^d x \left(1 + dm^2\rho^2 + (\nabla\rho)^2\right) \tag{4.13}$$

where ∇ stands for a spatial derivative.

Varying the area with respect to ρ, the RT surface obeys the equation

$$\nabla^2 \rho - dm^2 \rho = 0. \tag{4.14}$$

The solutions to this equation grow or decay as $e^{\pm\sqrt{dm^2}|\vec{y}|}$. The growing solution is the physical one, corresponding to the RT surface heading away from the horizon. Given an initial value $\rho_{\min}$, after a distance $|\vec{y}| \sim \frac{\log \rho_{\min}}{\sqrt{dm^2}}$, the RT surface is an order one distance from the horizon and the small ρ expansion is invalid. After it has significantly separated from the horizon, the RT surface is expected to reach the boundary after a further distance in $\vec{y}$ that is roughly independent of its extent on the horizon. Hence, for the purposes of determining the minimum size boundary region needed to include a point $\rho(t)$ near the horizon, it suffices to account just for the extent near the horizon since the minimum size boundary region is large but the part away from the horizon doesn't give a large contribution.

Since $\rho(t)$ is exponentially small for large t, setting $\rho_{\min} = \rho(t)$ gives a way to determine the minimum extent of A of balancing the exponential growth of the RT surface with the exponentially small initial condition. As discussed in the previous paragraph, taking the radius r_A of the disk A to be roughly the value of $|\vec{x}|$ (not $\vec{y}$) needed to escape the near horizon region, the balance is

$$\sqrt{dm^2}\, \frac{\ell u_h^{\theta/d}}{u_h u_0^{\theta/d}} r_A = \frac{2\pi t}{\beta}. \tag{4.15}$$

The right hand side can be simplified to give

$$\sqrt{\frac{(d-\theta)(d-\theta+z)}{2}}\, \frac{r_A}{u_h} = \frac{2\pi t}{\beta}, \tag{4.16}$$

which gives ballistic expansion with a velocity

$$\frac{r_A}{t} = \frac{2\pi u_h}{\beta} \sqrt{\frac{2}{(d-\theta)(d-\theta+z)}}. \tag{4.17}$$

Converting u_h to a temperature gives

$$\frac{r_A}{t} = \sqrt{\frac{8\pi^2}{(d-\theta)(d-\theta+z)}} \left(\frac{d-\theta+z}{4\pi}\right)^{1/z} \frac{\left(u_0^{z-1}\beta\right)^{1/z}}{\beta}. \tag{4.18}$$

Comparing to the holographic OTOC calculation, the rate of growth of the entanglement wedge is indeed the given by the butterfly velocity, which for these geometries is also

$$v_B = \sqrt{\frac{8\pi^2}{(d-\theta)(d-\theta+z)}} \left(\frac{d-\theta+z}{4\pi}\right)^{1/z} \frac{\left(u_0^{z-1}\beta\right)^{1/z}}{\beta}. \tag{4.19}$$

In the limit $\theta \to 0$ and $z \to 1$, which is the CFT limit, the butterfly velocity is

$$v_B \to \sqrt{\frac{d+1}{2d}}. \tag{4.20}$$

This is the value reported in the main text. It is interesting to compare this to the speed of sound, which in a CFT is $v_s = \sqrt{\frac{1}{d}}$ and which obeys $v_s \leq v_B$.

E   Entanglement growth bound

In this appendix we present a more detailed argument, following the one sketched at the end of Sec. 2, for $v_E \leq v_B$. The idea, again, is that the region A_t contains all the information in A if the velocity defining A_t is the butterfly velocity (plus a buffer that grows that takes care of details like the broadened wavefront discussed in Sec. 3). In terms of relative entropy, we expect something like $S(\rho_{A_t} \| \sigma_{A_t}) \geq S(\rho_A \| \sigma_A)$ where σ is a thermal state. Such a statement can then be used to show $v_E \leq v_B$. For convenience, I have highlighted all the non-trivial assumptions made in the argument. Most are technical and can be shown to hold in some restricted settings.

The system is d-dimensional and defined by a Hamiltonian which is a sum of local terms

$$H = \sum_x h_x. \tag{5.1}$$

The local Hilbert space dimension, $\dim \mathcal{V}_r$, is assumed to be finite, although it could be large. We consider a set Ω of uniform states which have a given energy density; we also include small perturbations by local operators in Ω provided they do not substantially change the energy density. The light cone physics is instantiated as follows: For any two states $|\psi_1\rangle$ and $|\psi_2\rangle$ in Ω, the time evolution operator obeys

$$|\langle\psi_1|U|\psi_2\rangle - \langle\psi_1|U_{A_t}U_{A^c}|\psi_2\rangle \leq \epsilon, \tag{5.2}$$

where for any region A, A_t is the expansion of A by $\delta\ell = v_L t$ in every direction. Note that we expect based on App. C that we can take $v_L = v_B$ (up to a buffer region) [**Assumption 1**], where v_B is obtained from OTOCs in a thermal state with the same energy density as states in Ω. For convenience in the argument, we will set $v_L = v_B + \delta v$, so that the result will be $v_E \leq v_B + \delta v$ for any $\delta v > 0$.

The linear size of A is denoted ℓ and we work in the regime where $\ell \gg \delta\ell$. Since A_t is the expansion of A by $\delta\ell = v_L t + \ell_0$, it overlaps with A^c. Using the form in Eq. 3.3, we expect that the squared commutator evaluated at $r = v_L t + \ell_0$ is

$$C(v_B t + \delta v t + \ell_0, t) \sim \exp\left(-\lambda \frac{(\delta v t)^{1+p}}{v_B (v_B t)^p}\right) = \exp\left(-\lambda \frac{(\delta v)^{1+p}}{v_B (v_B t)^p} t\right). \tag{5.3}$$

Hence for any $\delta v > 0$, the OTOC decays at least exponentially in time. We therefore expect, following App. C, that ϵ may be taken to decay exponentially with time [**Assumption 2**].

Let $\tilde{U} = U_{A'}U_{A^c}$ and let $\rho_0 = |\psi_0\rangle\langle\psi_0|$ be the initial state of the whole system. For convenience in the argument, ρ_0 is taken to be pure. We also assume, following the analysis in App. C, that $\tilde{U}$ does not substantially increase the energy density, so that both $U|\psi\rangle$ and $\tilde{U}|\psi\rangle$ are in Ω for all $|\psi\rangle \in \Omega$ [**Assumption 3**].

We next compute the trace distance between $U\rho_0 U^\dagger$ and $\tilde{U}\rho_0\tilde{U}^\dagger$. Since these states are pure, it follows that

$$\|U\rho_0 U^\dagger - \tilde{U}\rho_0\tilde{U}^\dagger\|_1 = 2 - 2|\langle\psi_0|U^\dagger\tilde{U}|\psi_0\rangle|^2. \tag{5.4}$$

Since $\tilde{U}|\psi_0\rangle$ and $U|\psi_0\rangle$ are in Ω, the inner product can be bounded as

$$|\langle\psi_0|U^\dagger\tilde{U}|\psi_0\rangle - 1| \leq \epsilon. \tag{5.5}$$

Hence the trace distance is bounded by

$$\|U\rho_0 U^\dagger - \tilde{U}\rho_0\tilde{U}^\dagger\|_1 \leq 4\epsilon. \tag{5.6}$$

To set up the relative entropy calculation, define $\sigma_0 = e^{-H/T}/Z$ to be the thermal state of the whole system at temperature T. The temperature is chosen to match the energy density of state ρ_0. The thermal state is independent of time,

$$\sigma = U\sigma_0 U^\dagger = \sigma_0, \tag{5.7}$$

and it is contained in Ω. It is also useful to define $\tilde{\sigma} = \tilde{U}\sigma_0\tilde{U}^\dagger$. Because $\tilde{U}$ is not the exact time evolution, the state $\tilde{\sigma}$ is not exactly time-independent and is not exactly equal to $\sigma = \sigma_0$, but it is close almost everywhere. Defining $\tilde{H}$ via the equation

$$\tilde{\sigma} = \frac{e^{-\tilde{H}/T}}{\tilde{Z}}, \tag{5.8}$$

we assume that $\tilde{H}$ is close to H with the difference concentrated near ∂A_t [**Assumption 4**]

The tilde variables are useful because the approximate time evolution operator $\tilde{U}$ can be reversed on A_t to reveal the original state on A. Defining $\tilde{\rho} = \tilde{U}\rho_0\tilde{U}^\dagger$ and recalling that $\tilde{U} = U_{A_t}U_{A^c}$, unitary invariance and monotonicity of relative entropy give

$$S(\tilde{\rho}_{A_t}\|\tilde{\sigma}_{A_t}) \geq S(\rho_{0,A}\|\sigma_{0,A}). \tag{5.9}$$

To use this equation, we need to know the physical content of $-\log\tilde{\sigma}_{A'}$ and $-\log\sigma_{0,A}$.

The operator $-\log\sigma_{0,A}$ is assumed to be approximately equal to $H_A/T +$ constant, the physical Hamiltonian truncated to region A divided by the temperature. The difference between H_A/T and $-\log\sigma_{0,A}$ is localized near the

boundary of A [**Assumption 5**]. We write

$$\sigma_{0,A} = \frac{e^{-(H_A + \delta H_A)/T}}{Z_A} \qquad (5.10)$$

where $Z_A = \text{Tr}(e^{-(H_A + \delta H_A)/T})$.

Now consider the relative entropy of $\rho_{0,A}$ relative to $\sigma_{0,A}$. We have

$$S(\rho_{0,A} \| \sigma_{0,A}) = \text{Tr}(\rho_{0,A}(H_A/T + \delta H_A/T)) + \log Z_A - S(\rho_{0,A}). \qquad (5.11)$$

Using the thermodynamic identity

$$\log Z_A = S(\sigma_{0,A}) - \text{Tr}(\sigma_{0,A}(H_A/T + \delta H_A/T)), \qquad (5.12)$$

the relative entropy is

$$\langle (H_A/T + \delta H_A/T) \rangle_{\rho_{0,A}} - \langle (H_A/T + \delta H_A/T) \rangle_{\sigma_{0,A}} - S(\rho_{0,A}) + S(\sigma_{0,A}). \quad (5.13)$$

Recall that T was chosen so that the thermal state had the same energy density as state $\rho_{0,A}$. Hence the average of H_A/T approximately cancels between $\sigma_{0,A}$ and $\rho_{0,A}$. Similarly, we expect that $S(\sigma_{0,A})$ will be approximately equal to $s_T|A|$ where s_T is the thermal entropy density provided A is large [**Assumption 6**]. These remainder terms can be combined with the δH_A terms to write

$$S(\rho_{0,A} \| \sigma_{0,A}) = s_T|A| - S(\rho_{0,A}) + \delta s \xi |\partial A| \qquad (5.14)$$

where ξ is some length scale that does not depend on the details of $\rho_{0,A}$ and δs is an entropy density.

The physics is that the various remainder terms in the relative entropy scale like the surface area $|\partial A| \sim \ell^{d-1}$ times a thickness ξ up to some constant δs which measures the number of relevant degrees of freedom. For lattice models that are well described at long wavelengths by a quantum field theory, ξ might be of order the thermal length and δs might be of order the thermal entropy. In this case, we would also expect that $\xi \delta s$ is not UV divergent, but for the purposes of the argument this is actually not crucial.

Now we apply the same logic to $S(\tilde{\rho}_{A_t} \| \tilde{\sigma}_{A_t})$. Since H and $\tilde{H}$ are close, the extensive term $\langle \tilde{H}_{A_t} \rangle$ still approximately cancels between $\tilde{\rho}_{A_t}$ and $\tilde{\sigma}_{A_t}$. The ϵ suppressed corrections can be encapsulated by writing

$$S(\tilde{\rho}_{A_t} \| \tilde{\sigma}_{A_t}) = s_T|A_t| - S(\tilde{\rho}_{A_t}) + \delta\tilde{s}\tilde{\xi}|\partial A_t| + O(\epsilon|A_t|). \qquad (5.15)$$

As indicated above, we will be taking $\delta\ell$ large enough so that $\epsilon|A_t|$ is small, even if the prefactor is UV sensitive. We can make $\epsilon|A_t|$ small while maintaining the limit $\delta\ell = v_L t \ll \ell$ because t need only be at least $d \log \ell + O(\ell^0)$ in size to make $\epsilon|A_t|$ small.

Assembling all the pieces, the monotonicity statement Eq. (5.9) translates to

$$s_T|A_t| - S(\tilde{\rho}_{A_t}) + \delta\tilde{s}\tilde{\xi}|\partial A_t| + O(\epsilon|A_t|) \geq s_T|A| - S(\rho_{0,A}) + \delta s\xi|\partial A|. \quad (5.16)$$

This can be rearranged to give

$$s_T(|A_t| - |A|) + \delta\tilde{s}\tilde{\xi}|\partial A_t| - \delta s\xi|\partial A| + O(\epsilon|A_t|) \geq S(\tilde{\rho}_{A_t}) - S(\rho_{0,A}). \quad (5.17)$$

In a translation invariant state, one can show that the entropy $S(\rho_{0,A})$ is an increasing function of the region size up to half the system size. Using $S(\rho_{A_t}) \geq S(\rho_A)$ gives

$$s_T(|A'| - |A|) + \delta\tilde{s}\tilde{\xi}|\partial A'| - \delta s\xi|\partial A| + O(\epsilon|A'|) \geq S(\tilde{\rho}_A) - S(\rho_{0,A'}). \quad (5.18)$$

The right hand side is the entropy of region A at time t under the $\tilde{U}$ dynamics minus the entropy of region A at time zero. Since we want the entropy growth under U instead of $\tilde{U}$, we use the Fannes-Audenaert inequality to write

$$|S(\tilde{\rho}_A) - S(\rho_A)| \leq 4\epsilon|A|\log\dim\mathcal{V}_r. \quad (5.19)$$

Absorbing the remainder $4\epsilon|A|\log\dim\mathcal{V}_r$ into the $O(\epsilon|A_t|)$ term gives

$$s_T(|A_t| - |A|) + \delta\tilde{s}\tilde{\xi}|\partial A'| - \delta s\xi|\partial A| + O(\epsilon|A_t|) \geq S(\rho_{A'}) - S(\rho_{0,A'}). \quad (5.20)$$

Note that we did need finite local Hilbert space dimension, but the extra time we need to ensure that $\epsilon\log\dim\mathcal{V}_r$ is small only grows logarithmically.

By construction we have $|A_t| - |A| = |\partial A|\delta\ell[1 + O(\delta\ell/\ell)]$. Also, $|\partial A_t| = |\partial A|[1 + O(\delta\ell/\ell)]$. Now we divide both sides of Eq. (5.20) by t to give

$$s_T v_L |\partial A|\left(1 + O(\delta\ell/\ell) + \frac{\delta\tilde{s}\tilde{\xi} - \delta s\xi + O(\delta\ell/\ell, \epsilon|A_t|)}{s_T v_L t}\right) \geq \frac{S(\rho_A) - S(\rho_{0,A})}{t}. \quad (5.21)$$

By taking t large enough such that the term in parenthesis is close to one, we conclude that the average rate of change of the entropy of A is bounded as,

$$s_T v_L |\partial A| \geq \left.\frac{S(\rho_A) - S(\rho_{0,A})}{t}\right|_{t\text{ large}}. \quad (5.22)$$

To be precise, we need $v_L t$ much larger than $(\delta s/s_T)\xi$ and $(\delta\tilde{s}/s_T)\tilde{\xi}$ and we need t large enough so that $\epsilon|A_t|$ is small. This must be achieved while maintaining $v_L t \ll \ell$. Of these conditions, the requirement that $\epsilon|A_t$ is likely the most stringent. In particular, it involves microscopic data of the system like the local Hilbert space dimension. However, because ϵ decreases exponentially with t, taking t of order $d\log\ell + \log(\text{microscopic constants})$ should suffice to guarantee $\epsilon|A_t|$ is small. The hierarchy

$$\ell \gg v_L t \gg d\log\ell + \text{constant} \quad (5.23)$$

can always be achieved for sufficiently large ℓ. Hence for sufficiently large regions we can show that the average rate of entropy increase is bounded by $s_T v_L |\partial A|$.

Physically, we might be interested in the rate of increase of entropy in a low energy state of a lattice model whose low energy dynamics is approximated by a conformal field theory, for example. In this case, scaling symmetry dictates that v_E is independent of region size, hence we can use a very large region to prove the entropy growth bound and then be confident that it applies also to smaller regions. Thus $v_E \leq v_L$ in this case, as claimed.

It is worth commenting that this bound is much tighter than the naive lattice bound in which the microscopic Hamiltonian appears. For example, if a is the lattice constant and J is the typical interaction strength, then the microscopic bound on the rate of change of entropy scales like $\frac{J|\partial A| \log \dim \mathcal{V}_r}{a^{d-1}}$. The Lieb-Robinson speed will be of order $v_{LR} \sim Ja$, so the microscopic bound can be written $\frac{v_{LR}|\partial A| \log \dim \mathcal{V}_r}{a^d}$. The factor of $\log \dim \mathcal{V}_r / a^d$ is of order the thermal entropy density at infinite temperature. Hence the improved bound amounts to replacing $s_\infty \to s_T$ and $v_{LR} \to v_B$.

Let me also make a few comments concerning the main assumptions. Assumptions 1, 2, and 3 are all related to the decomposition of U into $U_{A_t} U_{A^c}$ based on App. C. The arguments there are physically plausible and careful but not entirely rigorous. Assumptions 4, 5, and 6 are technical assumptions about the structure of thermal states. They are also physically plausible, at least for chaotic systems, but there is work to be done is they are to be proved in general or from some simpler assumption.